La natura dello spazio e del tempo

Due specialisti della relatività espongono le loro differenti opinioni sull'universo, sulla sua evoluzione e sulle conseguenze della meccanica quantistica

di Stephen W. Hawking e Roger Penrose

Nel 1994 Stephen W. Hawking e Roger Penrose tennero un ciclo di conferenze pubbliche sulla relatività generale presso l'Isaac Newton Institute for Mathematical Sciences dell'Università di Cambridge. Da queste conferenze, pubblicate in Italia da Sansoni col titolo *La natura dello spazio e del tempo,* abbiamo scelto alcuni brani dai quali bene si coglie la differenza tra le posizioni dei due scienziati. Nonostante le loro comuni radici scientifiche (Penrose fece parte della commissione di dottorato di Hawking a Cambridge), essi hanno una diversa visione della meccanica quantistica e delle sue conseguenze sull'evoluzione dell'universo. In particolare, Hawking e Penrose hanno opinioni contrastanti sul destino dell'informazione contenuta in un buco nero e sul motivo per cui la fine dell'universo sarà diversa dal suo inizio.

Una delle scoperte più importanti di Hawking, che risale al 1973, fu che i buchi neri emettono particelle a causa di effetti quantistici. Ciò fa sì che il buco nero evapori, sicché è possibile che della sua massa originale da ultimo non resti nulla. Ma durante la sua formazione un buco nero ingerisce un'enorme quantità di dati relativi al tipo, alle proprietà e alla configurazione delle particelle che vi cadono dentro. Benché la teoria quantistica imponga che queste informazioni si conservino, il loro destino ultimo è un punto molto controverso. Hawking e Penrose ritengono entrambi che, quando irradia, un buco nero perda le informazioni che conteneva, ma secondo Hawking la perdita è irreversibile, mentre Penrose sostiene che essa è compensata da misurazioni spontanee di stati quantistici che reintroducono informazioni nel sistema.

Ambedue gli scienziati ritengono che per descrivere la natura sia necessario formulare una teoria quantistica della gravitazione, ma hanno opinioni diverse su alcuni aspetti di questa teoria. Secondo Penrose, benché le forze fondamentali della fisica delle particelle siano simmetriche rispetto al tempo (cioè non mutino se il segno del tempo viene invertito), la gravità quantistica violerà questa simmetria. L'asimmetria temporale consentirà quindi di spiegare perché all'inizio l'universo era estremamente uniforme, come si deduce dalla radiazione di fondo a microonde residua dell'esplosione primordiale, mentre la fine dell'universo dovrà essere disordinata.

Penrose tenta di inglobare questa asimmetria temporale nella sua ipotesi della curvatura di Weyl. Come scoprì Einstein, la presenza della materia conferisce una curvatura allo spazio-tempo, il quale tuttavia può anche possedere una certa curvatura intrinseca, che viene chiamata curvatura di Weyl. Le onde gravitazionali e i buchi neri, per esempio, consentono allo spazio-tempo di curvarsi anche in regioni vuote. Nell'universo primordiale probabilmente la curvatura di Weyl era nulla, ma, secondo Penrose, quando l'universo si avvierà verso la morte, il gran numero di buchi neri produrrà un'accentuata curvatura di Weyl, e questa circostanza distinguerà la fine dell'universo dal suo inizio.

Hawking è d'accordo che la «grande contrazione» finale sarà diversa dal big bang, ma non accetta che nelle leggi di natura vi sia un'asimmetria temporale. Secondo lui, la ragione soggiacente della differenza sta nel modo in cui è programmata l'evoluzione dell'universo: egli postula una sorta di democrazia, per cui nell'universo non esisterebbero punti privilegiati. Quindi l'universo non può avere un confine. A suo parere, quest'assenza di confini spiega l'uniformità della radiazione di fondo a microonde.

Da ultimo, i due fisici danno interpretazioni diverse della meccanica quantistica. Hawking ritiene che a una teoria si debba chiedere solo di fornire previsioni in accordo coi dati, mentre Penrose crede che per spiegare la realtà non basti confrontare le previsioni con i risultati sperimentali e sottolinea che la teoria quantistica impone la «sovrapposizione» delle funzioni d'onda, da cui possono derivare conseguenze assurde. I due scienziati riprendono così i temi della famosa controversia tra Einstein e Niels Bohr sulla bizzarra natura della meccanica quantistica.

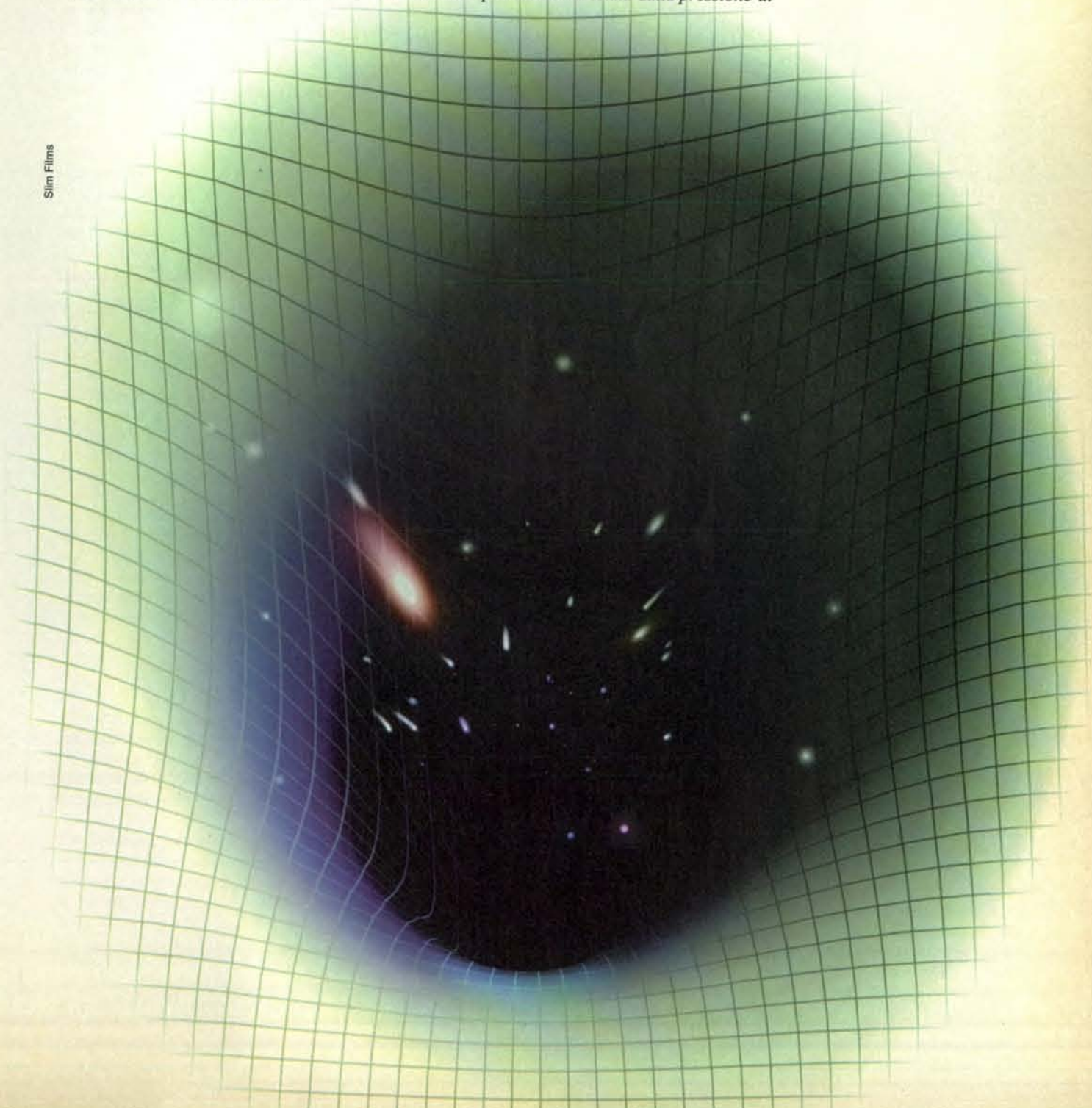
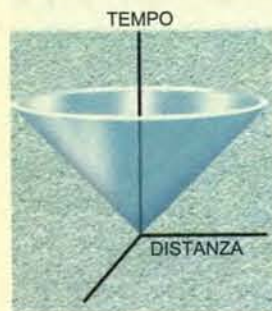
La redazione

Stephen Hawking sui buchi neri quantistici

La teoria quantistica dei buchi neri... sembra condurre a un nuovo livello di imprevedibilità in fisica, oltre all'indeterminazione abitualmente associata alla meccanica quantistica. Ciò si deve al fatto che i buchi neri sembrano avere un'entropia intrinseca e perdere informazione dalla nostra regione dell'universo. Dovrei dire che queste sono tesi controverse: molte persone che lavorano sulla gravità quantistica, compresi quasi tutti coloro che sono entrati in questo campo provenendo dalla fisica delle particelle, rifiuterebbero istintivamente l'idea che si possa perdere informazione sullo stato quantico di un sistema. Essi hanno avuto però ben poco successo nei loro tentativi di mostrare come si possano estrarre informazioni da un buco nero. Io credo che saranno infine costretti ad accettare il mio suggerimento che l'informazione è andata perduta, così come sono stati costretti ad ammettere che i buchi neri irraggiano, cosa che era contraria a tutti i loro preconcetti...

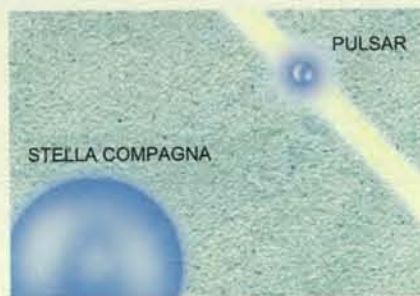
Il fatto che la gravità sia attrattiva significa che tenderà a formare concentrazioni di materia, le quali daranno origine a oggetti come stelle e galassie. Queste possono resistere per un certo tempo alla tendenza a un'ulteriore contrazione grazie alla pressione termica nel caso delle stelle, o alla rotazione e a moti interni nel caso di galassie. Col tempo, però, il calore o il momento angolare si dissipano e l'oggetto comincia a contrarsi. Se la massa è inferiore a una volta e mezzo circa la massa del Sole, la contrazione può essere arrestata dalla *pressione di*

PRESSIONE DI DEGENERAZIONE
Due elettroni o due neutroni non possono occupare il medesimo stato quantico. Pertanto, quando un insieme di queste particelle viene compresso in un piccolo volume, quelle che occupano gli stati quantici più alti acquistano un'energia elevata. Allora il sistema si oppone a un'ulteriore compressione esercitando una spinta verso l'esterno detta pressione di degenerazione.

Slim Films

CONI DI LUCE
Per raffigurare lo spazio-tempo, i fisici riportano di solito il tempo su un asse verticale e lo spazio su un asse orizzontale. Nello schema, i raggi di luce uscenti da un arbitrario punto dello spazio si propagano lungo la superficie di un cono verticale. Dato che nessun segnale fisico si può propagare con velocità superiore a quella della luce, tutti i segnali che escono dal punto risultano confinati all'interno del cono di luce.

SUPERFICIE NULLA
Una superficie dello spazio lungo la quale si propaga la luce è detta superficie nulla. La superficie nulla che circonda un buco nero, chiamata orizzonte degli eventi, è sferica. Ciò che cade entro un orizzonte degli eventi non può più uscirne.

MOMENTI DI MULTIPOLO
La dinamica di un oggetto può essere riassunta determinandone i momenti di multipolo. Ciascun momento viene calcolato dividendo l'oggetto in minuscoli elementi, moltiplicando una o più volte la massa di ciascun elemento per la sua distanza dal centro e poi sommando questi prodotti per tutti gli elementi. Una sfera, per esempio, ha un momento di monopolo, mentre un manubrio da ginnastica ha un momento di dipolo, che gli consente di acquistare con facilità un momento angolare.

PULSAR
Alcune stelle, morendo, subiscono un collasso e divengono stelle di neutroni, oggetti massicci costituiti solo di neutroni densamente ammassati. Le stelle di neutroni che ruotano velocemente diventano pulsar, così chiamate perché emettono impulsi di radiazione elettromagnetica a intervalli di millisecondi con estrema regolarità. Talvolta una pulsar orbita intorno a un'altra stella di neutroni, formando in questo modo una stella doppia.

Illustrazioni di Laurie Grace

degenerazione di elettroni o neutroni. L'oggetto si stabilizzerà nella forma, rispettivamente, di una nana bianca o di una stella di neutroni. Se invece la massa è superiore a questo limite, non c'è nulla che possa arrestarne la contrazione e impedirle di continuare a contrarsi. Una volta che il volume di quest'oggetto sia diminuito al di sotto di una certa grandezza critica, il campo gravitazionale alla sua superficie sarà così intenso che i *coni di luce* saranno orientati verso l'interno... Potete vedere che persino i raggi che riescono a uscire sono inclinati l'uno verso l'altro e sono perciò convergenti anziché divergenti. Ciò significa che c'è una superficie intrappolata chiusa...

Dev'esserci quindi una regione dello spazio-tempo da cui non è possibile evadere all'infinito. Questa regione viene detta buco nero. Il suo confine è chiamato l'orizzonte degli eventi ed è una *superficie nulla* formata dai raggi di luce che non riescono per un'inezia a sfuggire verso l'esterno...

...quando un corpo collassa a formare un buco nero, va perduta una grande quantità di informazione. Il corpo in preda al collasso gravitazionale è descritto da un numero di parametri molto grande. Ci sono i tipi di materia e i *momenti multipolari* della distribuzione di massa. Eppure il buco nero che si forma è del tutto indipendente dal tipo di materia e perde rapidamente tutti i momenti multipolari tranne i primi due: il momento di monopolo, che è la massa, e il momento di dipolo, che è il momento angolare.

Questa perdita d'informazione non aveva in realtà molta importanza nella teoria classica. Si potrebbe dire che tutta l'informazione sul corpo collassante era ancora contenuta all'interno del buco nero. Sarebbe molto difficile per un osservatore esterno al buco nero determinare com'era il corpo che sta subendo il collasso gravitazionale. Nella teoria classica questo compito era però ancora possibile in linea di principio. L'osservatore non perdeva mai realmente di vista il corpo collassante. Questo gli sembrava rallentare sempre più il suo movimento e diventare sempre meno luminoso man mano che si avvicinava all'orizzonte degli eventi, ma l'osservatore riusciva ancora a vedere di che cosa era fatto e com'era distribuita la sua massa.

La teoria quantistica cambiò però tutto questo. Innanzitutto il corpo in collasso emetteva - prima di attraversare l'orizzonte degli eventi - solo un numero limitato di fotoni, i quali erano del tutto insufficienti a trasportare tutta l'informazione sul corpo stesso. Ciò significa che, nella teoria quantistica, un osservatore esterno non ha alcun modo per misurare lo stato del corpo collassato. Si potrebbe anche non attribuire una grande importanza a questo fatto, dato che l'informazione sarebbe ancora contenuta nel buco nero, nonostante l'impossibilità di misurarla dall'esterno. Ma a questo punto interviene il secondo effetto della teoria quantistica dei buchi neri...

...la teoria quantistica comporta un'irradiazione e perdita di massa dei buchi neri. Pare che essi siano destinati a sparire infine completamente, portando con sé l'informazione contenuta al loro interno. Io addurrò argomenti a sostegno della tesi che quest'informazione va davvero perduta e che non ritorna sotto alcuna forma. Come mostrerò, questa perdita d'informazione introdurrà un nuovo livello di incertezza in fisica, in aggiunta all'indeterminazione abitualmente associata alla teoria quantistica. Purtroppo, diversamente dal principio di indeterminazione di Heisenberg, questo livello extra sarà piuttosto difficile da confermare sperimentalmente nel caso dei buchi neri.

Roger Penrose su teoria quantistica e spazio-tempo

Le grandi teorie fisiche del XX secolo sono state la teoria quantistica, la relatività ristretta, la relatività generale e la teoria quantistica dei campi. Queste teorie sono legate l'una all'altra: la relatività generale fu costruita sulla base della relatività ristretta e la teoria quantistica dei campi si fonda sulla relatività ristretta e sulla teoria quantistica.

Qualcuno ha sostenuto che la teoria quantistica dei campi sia la teoria fisica più esatta mai esistita, potendo vantare un margine di imprecisione di circa $1:10^{11}$. Vorrei però sottolineare che recentemente è stata verificata, in un certo senso chiaro, la precisione della relatività generale, la quale è risultata avere un margine d'errore non superiore a $1:10^{14}$ (e a quanto pare il limite dipende semplicemente dal grado di precisione degli orologi sulla Terra). Mi riferisco alla *pulsar* binaria di Hulse-Taylor PSR 1913 + 16, un paio di stelle di neutroni in orbita attorno a un baricentro comune, una delle quali è una pulsar. La relatività generale predice un graduale rimpicciolimento della sua orbita (e un conseguente lento accorciamento del suo periodo) a causa della perdita di energia attraverso l'emissione di onde gravitazionali. Questo fenomeno è stato effettivamente osservato, e l'intera descrizione del moto... concorda con la relatività generale (comprendente secondo me la teoria newtoniana), col margine d'errore notevolmente piccolo segnalato sopra su un periodo accumulato di vent'anni. Gli scopritori di questo sistema hanno ora giustamente ricevuto il premio Nobel per il loro lavoro. I teorici quantistici hanno sempre sostenuto che, in conseguenza della precisione della loro teoria, si dovrebbe modificare la relatività generale per adattarla a essa, ma oggi io penso che sia piuttosto la teoria quantistica dei campi a doversi sottoporre a qualche ritocco.

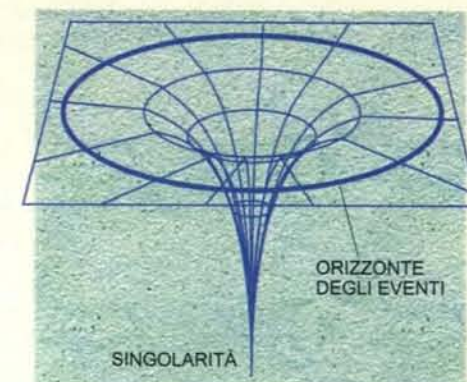
Benché le quattro teorie menzionate abbiano avuto un successo notevole, non sono senza problemi... La relatività generale predice l'esistenza di *singolarità* dello spazio-tempo. Nella teoria quantistica c'è il «problema della misurazione», di cui mi occuperò in seguito. Si può ritenere che la soluzione dei vari problemi che affliggono queste teorie si trovi nel fatto che, prese a sé, sono incomplete. Per esempio, molti prevedono che la teoria quantistica dei campi eliminerà in qualche modo le singolarità della relatività generale...

Vorrei ora parlare della perdita di informazione nei buchi neri, che sostengo sia pertinente a quest'ultimo problema. Sono d'accordo con quasi tutto ciò che Stephen ha da dire sull'argomento. Mentre però Stephen considera la perdita d'informazione dovuta ai buchi neri un'incertezza di più in fisica, al di là dell'indeterminazione della teoria quantistica, io la considero un'indeterminazione «complementare»... Può darsi che una piccola quantità d'informazione sfugga, al momento dell'evaporazione del buco nero... ma questo piccolo guadagno di informazione sarà molto minore della perdita d'informazione nel collasso (in quello che io considero un quadro ragionevole quanto qualsiasi altro della scomparsa finale del buco nero).

Se, per fare un esperimento concettuale, noi chiudiamo questo sistema in una grande scatola, possiamo considerare l'evoluzione dello *spazio delle fasi* della materia all'interno della scatola. Nella regione dello spazio delle fasi corrispondente a situazioni in cui è presente un buco nero, le traiettorie dell'evoluzione fisica convergeranno e i volumi che seguono queste traiettorie si contrarranno. Di questi sviluppi è responsabile la perdita d'informazione nella singolarità nel buco nero. Questa contrazione è in diretta contraddizione col teorema della meccanica classica ordinaria noto come teorema di Liouville, il quale dice che i volumi nello spazio delle fasi rimangono costanti... Lo spazio-tempo di un buco nero viola dunque questa conservazione. A mio modo di vedere, però, questa perdita di volume dello spazio delle fasi è controbilanciata da un processo di misurazione quantica «spontanea» in cui si guadagna informazione e i volumi dello spazio delle fasi aumentano. Ecco perché io considero l'indeterminazione dovuta alla perdita d'informazione nei buchi neri «complementare» all'indeterminazione nella teoria quantistica : l'una è il rovescio della medaglia dell'altra...

Per spiegare questo punto consideriamo l'esperimento concettuale del *gatto di Schrödinger*. L'esperimento, da me leggermente modificato, descrive la triste sorte di un gatto chiuso in una scatola, dove una sorgente di luce emette un fotone che incontra uno specchio semiargentato; questo scinde la funzione d'onda del fotone in due parti, una delle quali viene riflessa mentre l'altra viene lasciata passare dallo specchio verso un rivelatore; se questo capta il fotone, aziona una pistola che uccide il gatto, mentre se non lo capta la pistola non spara e il gatto rimane vivo e vegeto. (So che Stephen non accetta che si maltrattino i gatti, neppure in un esperimento mentale!) La funzione d'onda del sistema è una sovrapposizione di queste due possibilità... Ma perché la nostra percezione non ci permette di cogliere sovrapposizioni macroscopiche di stati come questi, e non solo le possibilità alternative macroscopiche «il gatto è morto» e «il gatto è vivo»?...

Vorrei suggerire che qualcosa va storto nelle sovrapposizioni delle geometrie alternative dello spa-

SINGOLARITÀ
Stando alla relatività generale, in certe condizioni estreme alcune regioni dello spazio-tempo acquistano una curvatura infinitamente grande e diventano singolarità: qui le usuali leggi fisiche cessano di valere. Nei buchi neri, per esempio, dovrebbero esservi singolarità nascoste dietro l'orizzonte degli eventi.

SPAZIO DELLE FASI
Un diagramma nello spazio delle fasi è un volume matematico a molte dimensioni che si costruisce assegnando una coordinata a ciascuna componente della distanza e del momento di ogni particella. Allora il moto di un insieme di particelle può essere rappresentato dal moto del corrispondente volume nello spazio delle fasi.

IL GATTO DI SCHRÖDINGER
Penrose invoca un esperimento ideale concepito da Einstein e usato poi da Erwin Schrödinger per studiare le complicazioni concettuali associate alle funzioni d'onda. Prima di una misura, si suppone che il sistema sia in una «sovrapposizione» di stati quantici o onde, e quindi il valore, per esempio, del momento è indeterminato. Dopo la misura, il valore della grandezza è noto e il sistema assume di colpo l'unico stato che corrisponde al risultato. Il significato della sovrapposizione iniziale e il procedimento con il quale il sistema «precipita» in un unico stato sono illustrati dal paradosso del gatto di Schrödinger.

zio-tempo che si verificherebbero quando comincia a essere coinvolta la relatività generale. Può darsi che una sovrapposizione di due geometrie diverse sia instabile, e che decada in una delle due possibilità alternative. Per esempio, le geometrie potrebbero essere gli spazi-tempi di un gatto vivo o di un gatto morto. Io chiamo questo decadere nell'una **o** nell'altra possibilità alternativa «riduzione obiettiva», che mi piace come nome perché ha un acronimo molto bello e appropriato, **OR** (che in inglese significa «o»). Quale relazione ha con ciò la lunghezza di Planck di 10^{-33} cm? Il criterio della natura per determinare quando due geometrie siano significativamente diverse sarebbe la *scala di Planck*, la quale fissa la scala di tempo alla quale si verifica la riduzione nelle possibilità alternative.

Hawking sulla cosmologia quantistica

Concluderò questa lezione occupandomi di un argomento su cui Roger e io abbiamo opinioni molto diverse: la freccia del tempo. C'è una distinzione molto chiara fra le direzioni in avanti e all'indietro del tempo nella nostra regione dell'universo. Per rendersi conto della differenza basta guardare un film proiettato all'indietro. Invece di vedere una tazzina che cade da un tavolo e si rompe, si vedrebbero cocci che si ricompongono in una tazzina integra, la quale salta sul tavolo. Che bello se la vita fosse così.

Le leggi locali a cui i campi fisici obbediscono sono simmetriche nel tempo o, più esattamente, sono *CPT-invarianti* (ossia invarianti rispetto alla coniugazione particella-antiparticella, alla riflessione spaziale, o parità, P, e alla riflessione temporale T). La differenza osservata fra il passato e il futuro deve quindi provenire dalle condizioni al contorno dell'universo. Supponiamo che l'universo sia spazialmente chiuso e che si espanda fino a raggiungere dimensioni massime, dopo di che comincerà a contrarsi. Come ha sottolineato Roger, l'universo sarà molto diverso ai due estremi di questa storia. In corrispondenza di quello che noi chiamiamo l'inizio dell'universo, pare che esso fosse molto omogeneo e regolare. Nella futura fase di collasso verso il Big Crunch ci attendiamo che esso diventi molto disordinato e irregolare. Poiché le configurazioni disordinate sono molto più numerose di quelle ordinate, ciò significa che le condizioni iniziali dovrebbero essere state incredibilmente precise.

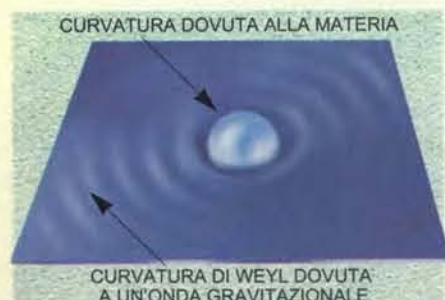
Pare, perciò, che ai due estremi del tempo debbano esserci condizioni al contorno molto diverse. Roger suggerisce che il *tensore di Weyl* dovrebbe svanire a un estremo del tempo ma non all'altro. Il tensore di Weyl è quella parte della curvatura dello spazio-tempo che non è determinata localmente dalla materia attraverso le equazioni di Einstein. Esso dovrebbe essere stato piccolo negli stati iniziali omogenei, ordinati, mentre dovrebbe essere grande nell'universo in collasso. Questa proposta distinguerebbe quindi le due estremità del tempo, e potrebbe perciò spiegare la freccia del tempo.

Io penso che la proposta di Roger possa essere considerata un'ipotesi di Weyl in più di un senso della parola. Innanzitutto, non è rispettata l'invarianza per la trasformazione CPT. Roger vede in questo fatto un pregio, ma io penso che ci si dovrebbe attenere alle simmetrie a meno che ci siano ragioni stringenti per rinunciare a esse. Come sosterrò, non è necessario rinunciare alla simmetria CPT. In secondo luogo, se nell'universo primitivo il tensore di Weyl fosse stato esattamente zero, l'universo sarebbe stato esattamente omogeneo e isotropo e lo sarebbe rimasto per sempre. L'ipotesi di Weyl proposta da Roger non potrebbe spiegare le fluttuazioni di fondo né le perturbazioni che hanno dato origine alle galassie e a noi stessi. Nonostante tutto questo, io penso che Roger abbia messo il dito su una differenza importante fra i due estremi del tempo. Il fatto però che a un estremo il tensore di Weyl fosse piccolo non dovrebbe essere imposto come una condizione al contorno ad hoc, ma dovrebbe essere dedotto da un principio più fondamentale, la *proposta dell'assenza di confine*...

Come possono essere diversi i due estremi del tempo? Perché le perturbazioni dovrebbero essere piccole a un estremo ma non all'altro? La ragione è che ci sono due possibili soluzioni complesse delle equazioni di campo...Ovviamente, una soluzione corrisponde a un estremo del tempo e l'altra all'altro... A un estremo l'universo era molto omogeneo e il tensore di Weyl era piccolissimo. Esso non poteva però essere esattamente zero in quanto in tal caso si sarebbe avuta una violazione del principio di indeterminazione. Dovevano invece essere presenti piccole fluttuazioni, che in seguito si amplificarono dando origine a galassie e a corpi come il nostro. Di contro l'universo dovrebbe essere molto irregolare e caotico all'altro estremo del tempo, con un tensore di Weyl tipicamente grande. Ciò spiegherebbe la freccia del tempo osservata e perché le tazzine cadano dai tavoli e si rompano sul pavimento invece di ricomporsi e saltare indietro sul tavolo.

SCALA DI PLANCK
La scala di Planck è una distanza irraggiungibilmente piccola (legata in meccanica quantistica a un intervallo di tempo impossibilmente breve e a una energia impossibilmente elevata) che compare se si combinano opportunamente le costanti fondamentali dell'attrazione gravitazionale, della velocità della luce e della meccanica quantistica. Essa rappresenta la distanza o l'energia alla quale gli attuali concetti di spazio, tempo e materia danno luogo alla gravità quantistica.

INVARIANZA CPT (CARICA-PARITÀ-TEMPO)
Questo principio impone che le teorie che descrivono le particelle restino valide anche quando la carica, la parità (o chiralità) e il tempo vengono tutti e tre invertiti. In altre parole, il comportamento di un elettrone con carica negativa e spin orario che si muova in avanti nel tempo deve coincidere con quello di un positrone con carica positiva e spin antiorario che si muova all'indietro nel tempo.

TENSORE DI WEYL
La curvatura dello spazio-tempo ha due componenti. Una deriva dalla presenza in esso della materia; l'altra, individuata dal matematico tedesco Hermann Weyl, si ha anche in assenza di materia. La grandezza matematica che descrive questa curvatura si chiama tensore di Weyl.

LA PROPOSTA DELL'ASSENZA DI CONFINE
Hawking ipotizza che l'evoluzione dell'universo sia spiegata dalla proposta dell'assenza di confine, formulata nel 1983 da lui e da James B. Hartle dell'Università della California a Santa Barbara. L'idea che l'universo non abbia confini impone vincoli alle soluzioni delle equazioni cosmologiche. Hawking ritiene che queste condizioni portino a una differenza tra l'inizio e la fine dell'universo, stabilendo così la direzione della freccia del tempo.

Penrose sulla cosmologia quantistica

Da quanto intendo della posizione di Stephen, non penso che il nostro disaccordo su questo punto sia molto grande (si veda l'*ipotesi della curvatura di Weyl*). Per una singolarità iniziale la curvatura di Weyl è prossima a zero... Stephen ha sostenuto che nello stato iniziale devono esserci piccole fluttuazioni, e ha quindi sottolineato l'irragionevolezza che la curvatura di Weyl iniziale fosse esattamente zero. Io non penso che su questo punto ci sia un vero disaccordo fra di noi. L'affermazione che la curvatura di Weyl sia zero alla singolarità iniziale è classica, e senza dubbio si deve ammettere una certa flessibilità circa la precisa formulazione dell'ipotesi. Piccole perturbazioni sono accettabili dal mio punto di vista, certamente nel regime quantistico. Abbiamo solo bisogno di qualcosa che mantenga tali perturbazioni molto vicine a zero...

Può darsi che la «proposta dell'assenza di confine» di Hartle e Hawking sia un buona candidata per la struttura dello stato iniziale. Mi pare però che noi abbiamo bisogno di qualcosa di molto diverso per occuparci dello stato finale. In particolare, una teoria che spiegasse la struttura delle singolarità dovrebbe violare T (riflessione temporale), PT (parità-riflessione temporale), CT (coniugazione di carica-riflessione temporale) e CPT (coniugazione di carica-parità-riflessione temporale) perché possa avere origine qualcosa del genere dell'ipotesi della curvatura di Weyl. Questa violazione della simmetria temporale potrebbe essere del tutto sottile; essa dovrebbe essere implicita nelle regole di quella teoria che va oltre la meccanica quantistica.

Hawking su fisica e realtà

Queste lezioni hanno mostrato in modo molto chiaro la differenza fra Roger e me. Roger è un platonico e io sono un positivista. Egli si preoccupa che il gatto di Schrödinger si trovi in uno stato quantico, che sia mezzo vivo e mezzo morto, e pensa che questa situazione non possa corrispondere alla realtà. A me questa cosa non dà alcun pensiero. Io non chiedo che una teoria corrisponda alla realtà perché non so quale sia questa realtà. La realtà non è una qualità che si possa verificare con una cartina al tornasole. Quel che mi interessa è che la teoria predica i risultati di misurazioni. La teoria quantistica lo fa con molto successo... Roger pensa che... il collasso della funzione d'onda introduca la violazione del prodotto CPT in fisica. Egli vede all'opera tale violazione in almeno due situazioni: la cosmologia e i buchi neri. Concordo con lui che potremmo introdurre l'asimmetria temporale nel modo in cui poniamo domande sulle osservazioni, ma rifiuto totalmente l'idea che ci sia qualche processo fisico che corrisponda alla riduzione della funzione d'onda, o che questo abbia qualcosa a che fare con la gravità quantistica o con la coscienza. Questa mi sembra magia, non scienza.

Penrose su fisica e realtà

La meccanica quantistica esiste da soli settantacinque anni. Non è molto al confronto, per esempio, con la teoria della gravità di Newton. Perciò non mi sorprenderebbe se dovesse essere modificata per oggetti molto macroscopici. All'inizio di questa discussione Stephen ha detto di ritenersi un positivista, mentre io sarei un platonico. Io sono lieto di sapere che lui è un positivista, ma penso che il punto cruciale qui sia piuttosto che io sono un realista. Inoltre, se si confronta questa discussione con quella famosa fra Bohr e Einstein, avvenuta una settantina di anni fa, io penserei che sia Stephen a interpretare la parte di Bohr, e io quella di Einstein. Einstein sosteneva infatti che deve esistere un mondo reale, non rappresentato necessariamente da una funzione d'onda, mentre Bohr sottolineava che la funzione d'onda non descrive un micromondo «reale» ma solo una «conoscenza» utile per fare previsioni.

È opinione diffusa che la controversia sia stata vinta da Bohr. In effetti, secondo la recente biografia di Einstein, *Einstein è vissuto qui*, di Abraham Pais, Einstein non ottenne più grandi risultati dal 1925 in avanti. È vero che egli non fece più grandi progressi, anche se le sue critiche penetranti furono sempre utili. Io credo che la ragione per cui Einstein non continuò a fare grandi progressi nella teoria quantistica fu che in essa mancava un ingrediente cruciale: la scoperta della radiazione dei buchi neri, fatta cinquant'anni dopo da Stephen. È questa perdita d'informazione, connessa con la radiazione dei buchi neri, a fornire la nuova svolta.

L'IPOTESI DELLA CURVATURA DI WEYL
Subito dopo il big bang l'universo ha una curvatura di Weyl piccola, mentre verso la fine del tempo la curvatura di Weyl è grande. Secondo Penrose è proprio questa curvatura che spiega la direzione della freccia del tempo.

NIELS BOHR

ALBERT EINSTEIN

Corbis-Bettmann; Laurie Grace

La fascia di Kuiper

Anziché terminare bruscamente con l'orbita di Plutone, il sistema solare contiene un'ampia regione occupata da piccoli corpi

di Jane X. Luu e David C. Jewitt

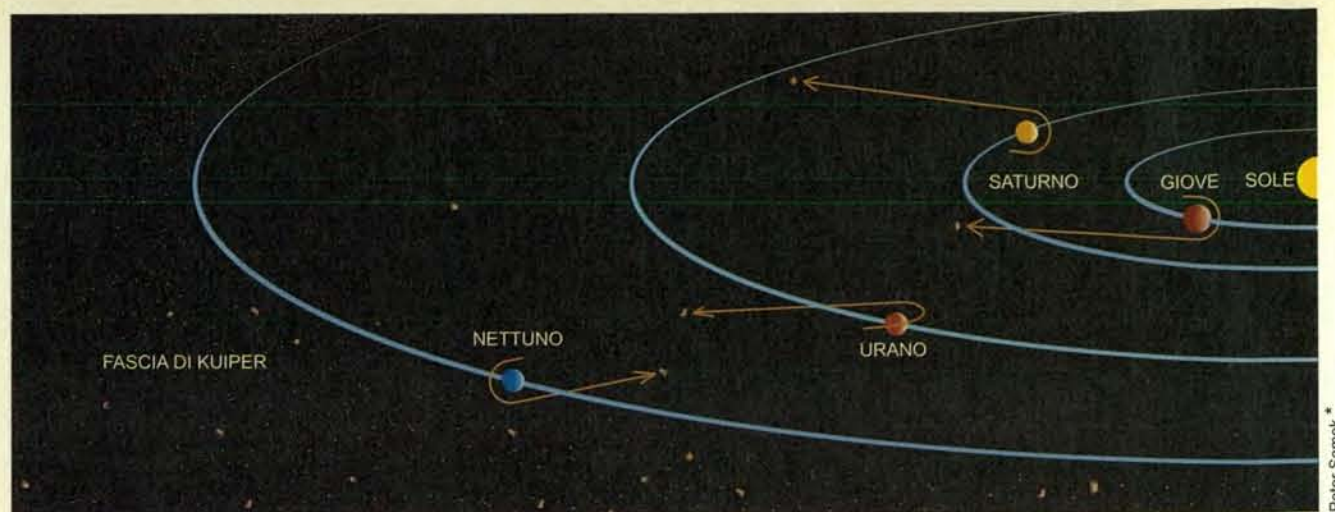
Dopo la scoperta di Plutone nel 1930, molti astronomi furono affascinati dalla possibilità di trovare un decimo pianeta in orbita intorno al Sole. Celato dalle enormi distanze dello spazio interplanetario, il misterioso «Pianeta X» poteva essere sfuggito all'osservazione anche dei migliori telescopi, o almeno così pensavano questi scienziati. Tuttavia sono passati decenni senza che venisse fatta alcuna scoperta, e gran parte dei ricercatori accetta ormai che il sistema solare sia limitato ai nove pianeti che ci sono familiari.

Molti scienziati, tuttavia, hanno cominciato a rivedere profondamente la loro concezione del sistema solare a partire dal 1992, quando il nostro gruppo identificò un piccolo corpo celeste - di poche centinaia di chilometri di diametro - situato a una distanza dal Sole superiore a quella di tutti i pianeti conosciuti. Da allora abbiamo identificato una trentina di oggetti simili nel sistema solare esterno. È probabile che vi sia un numero considerevole di questi piccoli corpi nella fascia di Kuiper, una regione così chiamata in onore di Gerard P. Kuiper, l'astronomo americano di origine olandese che nel 1951 propose per primo che il sistema solare potesse includere una simile famiglia di oggetti lontani.

Ma che cosa spinse Kuiper a ipotizzare, quasi un quarto di secolo fa, che il piano del sistema solare fosse popolato da numerosi piccoli corpi orbitanti a grande distanza dal Sole? Questa sua convinzione derivava dalla conoscenza del comportamento delle comete, agglomerati di roccia e ghiaccio la cui orbita le porta, a intervalli regolari, ad avvicinarsi al Sole provenendo dalle regioni esterne del sistema solare. Molti di questi oggetti relativamente piccoli assumono periodicamente un aspetto spettacolare, non appena la radiazione solare li riscalda a sufficienza da creare una grande chioma luminosa, fatta di polvere e gas, e una coda allungata.

Gli astronomi hanno compreso da lungo tempo che queste comete attive devono essere membri relativamente nuovi del sistema solare interno. Un corpo come la cometa di

Nelle zone remote del sistema solare si conserva materia primordiale residua dell'epoca di formazione dei pianeti. In questa fase primitiva Plutone (*in primo piano*) potrebbe avere catturato il suo satellite Caronte (*a destra*), allontanando nel contempo un terzo corpo nello spazio (*in alto*). All'epoca la regione doveva contenere grandi quantità di polvere, nonché molti oggetti della fascia di Kuiper in via di accrescimento.

Alfred T. Kamajian

Nel corso delle prime fasi dell'esistenza del sistema solare la forza di gravità esercitata dai pianeti sortì l'effetto di spazzare via i piccoli corpi che si trovavano all'interno dell'orbita di Nettuno. Alcuni di questi piccoli oggetti precipitarono in direzione del Sole, mentre altri si allontanarono verso la remota nube di Oort (*non mostrata nell'illustrazione*).

Halley, che passa nelle vicinanze del Sole ogni 76 anni, perde circa un decimillesimo della propria massa nel corso di ogni visita. Questa cometa sopravviverà quindi solo per circa 10 000 orbite, pari forse a mezzo milione di anni in tutto. Dato che simili comete ebbero origine durante la formazione del sistema solare, 4,5 miliardi di anni fa, ormai dovrebbero aver perduto totalmente i loro costituenti volatili, trasformandosi in nuclei rocciosi inattivi o in correnti rarefatte di polvere. Perché allora esistono ancora tante comete in grado di stupirci con le loro pirotecniche apparizioni?

Le comete attualmente attive si sono formate nelle prime fasi di vita del sistema solare, ma fino a tempi recenti si sono conservate in uno stato inattivo, per lo più all'interno di quel «congelatore» celeste che è la nube di Oort. L'astronomo olandese Jan H. Oort ipotizzò nel 1950 l'esistenza di questa sfera di materia cometaria. Egli riteneva che la nube avesse un diametro di circa 100 000 UA (unità astronomica, una distanza definita pari alla separazione media fra Terra e Sole, ossia circa 150 milioni di chilometri) e che contenesse centinaia di miliardi di comete. Secondo la concezione di Oort, gli effetti gravitazionali casuali dovuti al passaggio delle stelle vicine strapperebbero dalle loro orbite stabili alcune comete della nube e le spingerebbero gradualmente verso il Sole.

Per buona parte dell'ultimo mezzo secolo l'ipotesi di Oort ha spiegato adeguatamente dimensioni e orientazione delle traiettorie seguite dalle cosiddette comete di lungo periodo (quelle che impiegano più di 200 anni per completare un'orbita intorno al Sole). Le osservazioni astronomiche dimostrano che questi corpi giungono nella zona occupata dai pianeti provenendo da direzioni casuali, come ci si aspetterebbe se le comete avessero origine in una regione sferica come la nube di Oort. Viceversa, l'ipotesi di Oort non poteva spiegare le orbite delle comete di breve periodo, che normalmente sono più piccole e solo leggermente inclinate rispetto all'eclittica (il piano dell'orbita terrestre).

SOLE
NETTUNO
Peter Samek
100 UNITÀ ASTRONOMICHE

Gli oggetti che orbitano nella fascia di Kuiper, a grande distanza dal Sole, potrebbero essere in numero enorme, ma non tutti risultano visibili dalla Terra. I corpi (*cerchi*) che potrebbero essere individuati con il telescopio situato sul Mauna Kea, nelle Hawaii, si trovano tipicamente vicino al bordo interno della fascia, come si vede in questa simulazione della distribuzione della materia lontana.

La maggior parte degli astronomi riteneva che le comete di breve periodo percorressero in origine orbite immense, orientate in maniera casuale (come fanno oggi le comete di lungo periodo), e che fossero state portate nella loro attuale configurazione orbitale dagli effetti gravitazionali dei pianeti, soprattutto di Giove. Non tutti gli scienziati però erano d'accordo con questa ipotesi. Già nel 1949 Kenneth Essex Edgeworth, studioso non professionista irlandese (che non apparteneva ad alcun istituto di ricerca), scrisse una memoria nella quale proponeva che nel sistema solare esterno potesse esistere un anello piatto costituito da comete. Nel suo articolo del 1951, anche Kuiper parlò di una simile fascia di comete, ma senza fare riferimento al precedente lavoro di Edgeworth.

Kuiper, insieme con altri, sostenne che il disco del sistema solare non terminasse bruscamente con Nettuno e Plutone (ognuno dei quali, a seconda della sua posizione nell'orbita, può avere la palma del pianeta più lontano dal Sole); egli immaginò invece una fascia, situata fra Nettuno e Plutone, consistente di materia residua della formazione dei pianeti. La densità della materia in questa regione esterna sarebbe stata troppo bassa per consentire l'accrescimento di grandi pianeti, ma questo non impediva che vi si trovassero oggetti più piccoli, forse di dimensioni asteroidali. Essendo situati a grande distanza dal Sole, questi residui sparsi di materia primordiale presumibilmente hanno sempre avuto una temperatura superficiale bassa; sembrava quindi probabile che questi oggetti lontani fossero composti di ghiaccio d'acqua e di vari gas allo stato solido, una composizione che li rendeva molto simili (se non identici) ai nuclei di comete.

L'ipotesi di Kuiper languì fino agli anni settanta, quando Paul C. Joss del Massachusetts Institute of Technology cominciò a chiedersi se la gravità di Giove potesse davvero trasformare efficientemente comete di lungo periodo in comete di breve periodo. Egli fece notare che la probabilità di una cattura gravitazionale era così bassa da non potere in alcun modo spiegare il gran numero di comete di breve periodo oggi esistenti. Altri ricercatori, tuttavia, non riuscirono a confermare questo risultato, e l'idea che la nube di Oort desse origine a tutte le comete, sia di lungo sia di breve periodo, rimase la più accettata.

Ma Joss aveva gettato il seme del dubbio, e a poco a poco altri astronomi cominciarono a mettere in discussione la teoria consolidata. Nel 1980 Julio A. Fernández (allora al Max-Planck-Institut für Aeronomie di Katlenburg-Lindau) aveva, per esempio, effettuato calcoli che indicavano come le comete di breve periodo potessero provenire da una zona appena oltre l'orbita di Nettuno, come proposto da Kuiper. Nel 1988 Martin J. Duncan dell'Università di Toronto, Thomas Quinn e Scott D. Tremaine (entrambi del Canadian Institute for Theoretical Astrophysics) utilizzarono simulazioni al calcolatore per studiare il meccanismo della cattura di una cometa da parte dei pianeti giganti gassosi. Come Joss, essi scoprirono che il processo non funzionava molto bene, il che sollevava dubbi sulla verosimiglianza delle concezioni tradizionali riguardo all'origine delle comete di breve periodo. In effetti, i loro studi rivelarono un'ulteriore discrepanza: essi infatti fecero notare che le poche comete che potrebbero essere state espulse dalla nube di Oort a causa dell'attrazione gravitazionale dei pianeti giganti tendono a formare uno sciame sferico, mentre le orbite delle comete di breve periodo giacciono per lo più in un piano vicino a quello dell'eclittica.

Duncan, Quinn e Tremaine ne dedussero che le comete di breve periodo devono essere state catturate a partire da orbite che non erano molto inclinate rispetto all'eclittica: forse provenivano da una fascia appiattita di comete situa-

Una serie di immagini ottenute nel 1992 per mezzo di sensori a scorrimento di carica (CCD) ha rivelato chiaramente l'oggetto QB_1, appartenente alla fascia di Kuiper, in moto sullo sfondo delle stelle fisse (*al centro e in basso*). Queste due immagini coprono solo una piccola parte dell'inquadratura completa (*in alto a destra*) che si è dovuta analizzare prima che gli autori potessero identificare QB_1 (*frecce*) e determinarne l'orbita (*in alto a sinistra*).

ta nel sistema solare esterno. Tuttavia la loro cosiddetta ipotesi della fascia di Kuiper non mancava di suscitare perplessità. Per far sì che i calcoli fossero trattabili, essi avevano considerato valori di massa per i pianeti esterni anche 40 volte superiori a quelli reali (aumentando così l'intensità dell'attrazione gravitazionale e accelerando l'evoluzione orbitale che desideravano studiare). Altri astrofisici si chiedevano se questo «gioco di prestigio» computazionale non avesse potuto condurre a una conclusione errata.

Ancora prima che Duncan, Quinn e Tremaine pubblicassero il loro lavoro, ci eravamo domandati se il sistema solare esterno fosse davvero vuoto, oppure se potesse essere occupato da piccoli corpi mai osservati. Nel 1987 iniziammo una ricognizione telescopica proprio allo scopo di rispondere a questa domanda. Il nostro intento era quello di cercare eventuali oggetti presenti nel sistema solare esterno sfruttando la minuscola quantità di luce solare da essi riflessa a distanze così enormi. Sebbene nei nostri primi tentativi avessimo impiegato lastre fotografiche, decidemmo ben presto che le prospettive apparivano migliori con l'uso di un rivelatore elettronico (un dispositivo a scorrimento di carica o CCD) fissato a uno dei nostri telescopi più grandi.

Abbiamo condotto gran parte del nostro studio utilizzando il telescopio da 2,2 metri dell'Università delle Hawaii sul Mauna Kea. La nostra strategia era quella di impiegare con questo strumento una schiera di CCD per ottenere quattro esposizioni da 15 minuti, in sequenza, di una particolare zona di cielo. Ci siamo poi serviti del calcolatore per mostrare le immagini della sequenza in rapida successione; un oggetto che si muova leggermente nell'immagine, sullo sfondo delle stelle (che appaiono fisse), viene così rivelato come appartenente al sistema solare.

Per cinque anni abbiamo continuato la ricerca, ottenendo solo risultati negativi. Ma la tecnologia a nostra disposizione migliorava così rapidamente che era facile conservare l'entusiasmo (se non i finanziamenti) nella continua ricerca della nostra elusiva preda. Il 30 agosto 1992 stavamo eseguendo la terza di una sequenza di quattro esposizioni, e nello stesso tempo alternavamo rapidamente le prime due immagini sullo schermo del calcolatore. Notammo così che la posizione di una «stella» debole sembrava cambiare lievemente da una immagine all'altra. Entrambi restammo ammutoliti: lo spostamento non era certo vistoso, ma sembrava reale. Quando confrontammo le prime due immagini con la terza, ci rendemmo conto di aver senza dubbio trovato qualcosa di insolito. Il suo lento moto nel cielo indicava che l'oggetto che avevamo scoperto doveva trovarsi addirittura al di là dell'afelio della lontana orbita di Plutone. Avevamo il sospetto che l'oggetto misterioso potesse essere un asteroide vicino che si muoveva in parallelo con la Terra (fenomeno che potrebbe provocare a sua volta un lento moto apparente), ma ulteriori misurazioni portarono a escludere questa possibilità.

Osservammo di nuovo lo strano corpo nelle due notti successive e ne misurammo accuratamente posizione, luminosità e colore. Comunicammo poi questi dati a Brian G. Marsden, direttore del Central Bureau of Astronomical Telegrams della International Astronomical Union presso lo Smithsonian Astrophysical Observatory di Cambridge nel Massachusetts. I suoi calcoli indicarono che l'oggetto da noi scoperto orbitava realmente a grande distanza dal Sole: 40 UA, ossia poco meno di quanto avessimo ipotizzato inizialmente. Egli assegnò al corpo celeste appena scoperto un nome formale, anche se non particolarmente fantasioso, basato sulla data della scoperta: «1992 QB_1». (Noi preferivamo chiamarlo «Smiley», in onore della celebre spia dei romanzi di John Le Carré, ma il nome non trovò molti sostenitori nella serissima comunità degli astronomi.)

Le nostre osservazioni mostravano che la luce riflessa da QB_1 è decisamente più rossa della luce solare che lo illumina. Questa strana colorazione si ritrova in un solo altro oggetto del sistema solare: un asteroide (o cometa) peculiare, denominato 5145 Pholus. I planetologi attribuiscono il colore rosso di questo corpo alla presenza sulla sua superficie di materiale scuro ricco di carbonio. La somiglianza fra QB_1 e 5145 Pholus non fece che aumentare il nostro entusiasmo nei primi giorni dopo la scoperta. Forse l'oggetto che avevamo appena localizzato era ricoperto da un qualche tipo di materiale rossastro ricco di composti organici. Quanto era grande questo nuovo mondo rugginoso? Dalla prima serie di misurazioni, stimavamo che QB_1 avesse un diametro compreso tra 200 e 250 chilometri, ossia circa 15 volte le dimensioni del nucleo della cometa di Halley.

Alcuni astronomi dubitarono in un primo momento che la scoperta di QB_1 implicasse realmente l'esistenza di una popolazione di oggetti nel sistema solare esterno, come Kuiper e altri avevano ipotizzato. Tuttavia i dubbi cominciarono a venir meno quando scoprimmo, nel marzo 1993, un secondo oggetto, lontano dal Sole quanto QB_1, ma localizzato sul lato opposto del sistema solare. Negli ultimi tre anni diversi altri gruppi di ricerca si sono uniti ai nostri sforzi, e le scoperte non si sono fatte attendere: l'attuale elenco di oggetti situati nella fascia di Kuiper, oltre l'orbita di Nettuno, comprende 32 membri.

I componenti noti della fascia di Kuiper presentano svariate caratteristiche comuni. Per esempio, si trovano tutti oltre l'orbita di Nettuno, che quindi potrebbe definire il margine interno della fascia. Questi corpi celesti di recente scoperta si muovono su orbite solo leggermente inclinate rispetto all'eclittica, a sostegno dell'ipotesi che essi costituiscano una fascia piatta di comete. Ciascuno dei membri della fascia di Kuiper, inoltre, è milioni di volte più debole del limite di visibilità a occhio nudo. I 32 corpi individuati hanno un diametro variabile tra 100 e 400 chilometri: sono quindi notevolmente più piccoli sia di Plutone (che ha un diametro di circa 2300 chilometri) sia del suo satellite Caronte (circa 1100 chilometri di diametro).

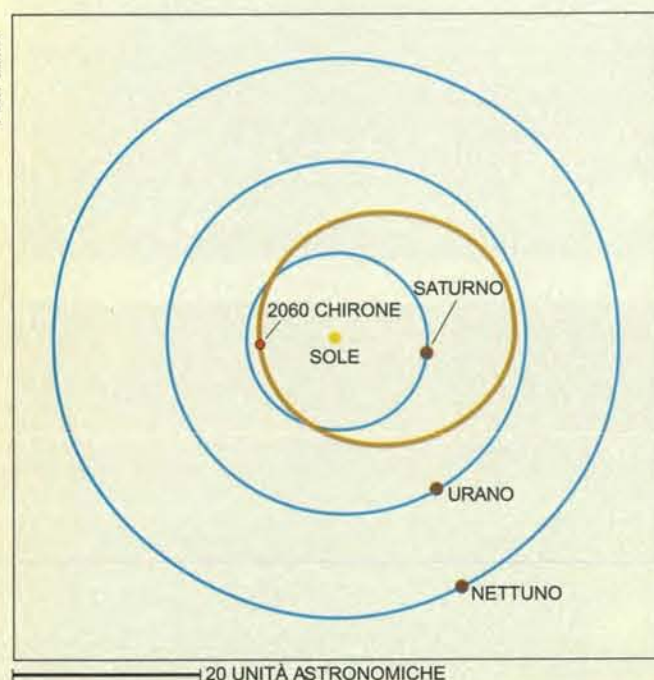
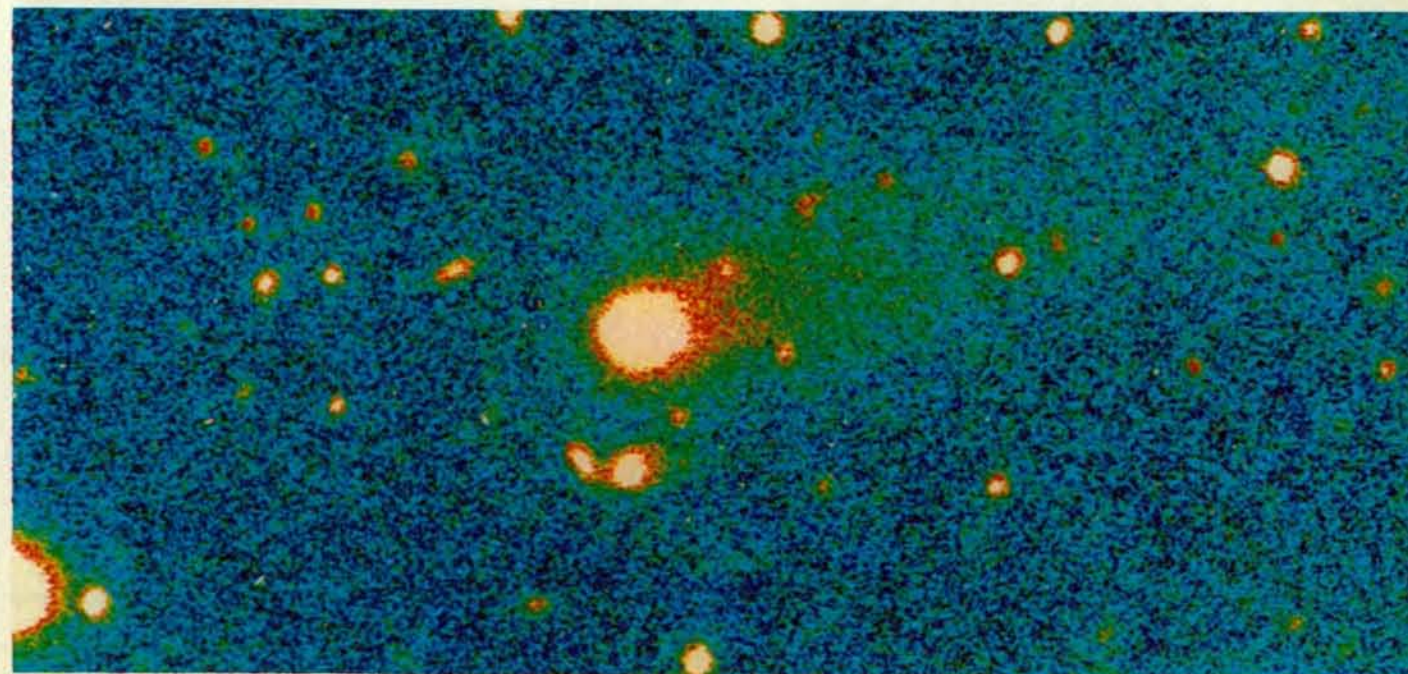
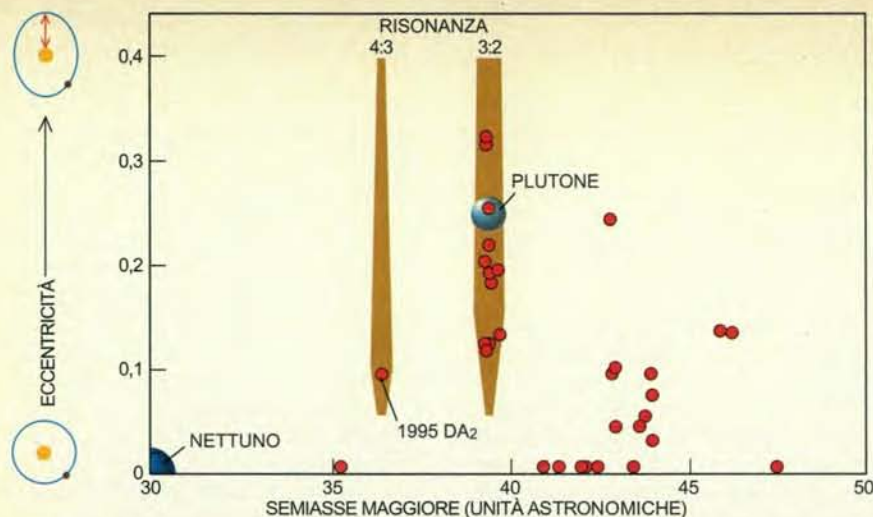
Il campione di oggetti finora identificati è abbastanza modesto, ma sufficiente per dimostrare al di là di ogni possibile dubbio l'esistenza della fascia di Kuiper. È anche chiaro che la popolazione totale della fascia deve essere considerevole; stimiamo che essa contenga almeno 35 000 oggetti di diametro superiore a 100 chilometri. Pertanto è probabile che la fascia di Kuiper abbia una massa complessiva centinaia di volte maggiore di quella della ben nota fascia degli asteroidi situata tra le orbite di Marte e di Giove.

Ammettiamo che la fascia di Kuiper sia ricca di materia; può davvero essere la sorgente delle comete di breve periodo, che si consumano così rapidamente? Matthew J. Holman e Jack L. Wisdom, che allora lavoravano entrambi al MIT, hanno affrontato questo problema servendosi di simulazioni al calcolatore. Essi hanno così dimostrato che, in un intervallo di 100 000 anni, l'influenza gravitazionale dei pianeti giganti gassosi (Giove, Saturno, Urano e Nettuno) allontana le comete che orbitano nelle loro vicinanze, espellendole nelle zone più esterne del sistema solare. Ma una frazione considerevole delle comete che orbitano oltre Nettuno può sfuggire a questo destino e rimanere nella posizione di partenza anche dopo 4,5 miliardi di anni. È dunque probabile che gli oggetti della fascia di Kuiper situati a più di 40 UA dal Sole si trovino in orbite stabili fin dalla formazione del sistema solare.

Gli astronomi ritengono inoltre che la massa contenuta nella fascia di Kuiper fosse sufficiente per formare tutte le comete di breve periodo che siano mai esistite; questa regione del sistema solare sembra quindi essere una buona candidata a «magazzino di comete».

Anche il meccanismo di trasferimento al di fuori della fascia oggi è ben conosciuto. Simulazioni al calcolatore hanno dimostrato che l'azione gravitazionale di Nettuno erode lentamente il bordo interno della fascia di Kuiper (la regione compresa entro 40 UA dal Sole), espellendo oggetti situati in questa zona verso il sistema solare interno. Molti di questi piccoli corpi finiscono per diventare comete che si consumano lentamente, ma alcuni - come la cometa Shoemaker-Levy 9, caduta su Giove nel luglio 1994 - possono trovare una fine violenta collidendo con un pianeta (o anche con il Sole). Altri possono essere intrappolati in una «fionda gravitazionale» che li scaglia verso le zone remote dello spazio interstellare.

Se la fascia di Kuiper è la fonte delle comete di breve periodo, si pone un'altra domanda ovvia: attualmente vi sono forse comete che hanno abbandonato questa regione e stanno dirigendosi verso il sistema solare interno? La risposta potrebbe essere costituita dai Centauri, un gruppo di oggetti che include il già menzionato 5145 Pholus. I Centauri seguono orbite immense che attraversano le orbite planetarie e sono intrinsecamente instabili; possono rimanere nella zona dei pianeti giganti solo per alcuni milioni di anni prima che le interazioni gravitazionali li scaglino fuori dal sistema solare o li trasferiscano in orbite più strette.

Dato che i loro tempi di permanenza in orbita sono di gran lunga più brevi dell'età del sistema solare, i Centauri non possono essersi formati dove si trovano attualmente. Tuttavia la natura delle loro orbite fa sì che risulti presso-

2060 Chirone potrebbe aver lasciato la fascia di Kuiper per immettersi nella sua attuale traiettoria che incrocia le orbite dei pianeti (*a sinistra*). Per quanto molto debole, il bagliore che circonda 2060 Chirone (*nella pagina a fronte*) indica che esso potrebbe essere correlato con altri corpi «attivi», come la cometa Peltier (*qui sopra*).

Fenomeni di risonanza regolano dimensione e forma delle orbite di molti oggetti della fascia di Kuiper. Un'orbita è descritta dall'eccentricità (ossia la deviazione dalla forma circolare) e dal semiasse maggiore (*freccette rosse*). Come Plutone, circa metà dei corpi conosciuti della fascia di Kuiper (*punti rossi*) compie due orbite intorno al Sole nello stesso tempo in cui Nettuno ne completa tre; si ha quindi una risonanza 3:2. L'oggetto 1995 DA_2 orbita in una delle altre possibili risonanze. Renu Malhotra del Lunar and Planetary Institute di Houston propone che questo fenomeno rispecchi le prime fasi di evoluzione del sistema solare, quando molti piccoli corpi furono espulsi nello spazio e i pianeti più grandi si allontanarono dal Sole. Durante questi movimenti verso l'esterno Nettuno potrebbe aver trascinato Plutone e numerosi altri corpi più piccoli nelle orbite risonanti che si osservano oggi.

Peter Samek

ché impossibile risalire con certezza al loro luogo di origine. Nonostante ciò, la fonte più vicina (e più probabile) è la fascia di Kuiper. I Centauri possono quindi essere «comete di transizione», ossia oggetti un tempo appartenenti alla fascia di Kuiper che si apprestano a iniziare una vita breve ma spettacolare nel sistema solare interno. Gli elementi più solidi a sostegno di questa ipotesi sono forniti da uno degli oggetti di questo gruppo: 2060 Chirone. Sebbene i suoi scopritori lo avessero inizialmente considerato solo un asteroide insolito, 2060 Chirone viene oggi risolutamente classificato come una cometa attiva dalla chioma debole ma persistente.

Via via che lo studio della fascia di Kuiper continua, alcuni astronomi hanno cominciato a chiedersi se da questo «serbatoio» potrebbe provenire qualcosa di più che semplici comete. È una coincidenza che Plutone, il suo satellite Caronte e Tritone, uno dei satelliti di Nettuno, si trovino tutti nelle vicinanze della fascia di Kuiper? Questa domanda è giustificata, in quanto Plutone, Caronte e Tritone hanno alcuni tratti in comune nelle loro proprietà fondamentali e differiscono drasticamente dai loro vicini.

La densità di Plutone e di Tritone, per esempio, è molto maggiore di quella di tutti i pianeti giganti gassosi del sistema solare esterno; inoltre anche il moto orbitale di questi corpi è alquanto strano. Tritone ruota intorno a Nettuno in senso «retrogrado», ossia opposto al senso orbitale di tutti i pianeti e della maggior parte dei satelliti. L'orbita di Plutone è fortemente inclinata rispetto all'eclittica e così allungata che incrocia addirittura l'orbita di Nettuno. Plutone tuttavia non rischia una collisione con il pianeta gigante in quanto le loro orbite hanno un particolare rapporto di risonanza, 3:2; ciò significa che, per ogni tre orbite di Nettuno intorno al Sole, Plutone ne compie due.

I pezzi di questo puzzle celeste possono essere ricomposti se si postula che Plutone, Caronte e Tritone siano gli ultimi superstiti di un gruppo, un tempo molto più grande, di oggetti di dimensioni simili. S. Alan Stern del Southwest Research Institute di Boulder fu il primo a proporre quest'idea nel 1991. I tre corpi potrebbero essere caduti sotto l'influenza di Nettuno, che catturò Tritone e bloccò Plutone - forse con Caronte a ruota - nella sua attuale risonanza orbitale.

È interessante il fatto che le risonanze orbitali sembrano influenzare anche la posizione di molti oggetti della fascia di Kuiper. Fino a metà dei corpi scoperti di recente sembrano avere lo stesso rapporto di risonanza 3:2 con Nettuno mostrato da Plutone; come quest'ultimo, essi potranno rimanere tranquillamente nelle loro orbite per miliardi di anni. (La risonanza impedisce a Nettuno di avvicinarsi troppo e perturbare l'orbita del corpo più piccolo.) Abbiamo battezzato questi oggetti della fascia di Kuiper «Plutini», ossia «piccoli Plutoni». A giudicare dalla limitata regione di cielo che abbiamo esaminato, stimiamo che debbano esistere diverse migliaia di Plutini di diametro superiore a 100 chilometri.

Le recenti scoperte di oggetti della fascia di Kuiper aprono nuove prospettive sul sistema solare esterno. Attualmente Plutone sembra avere un posto a parte solo perché è più grande di tutti gli altri membri della fascia di Kuiper, e si può persino dubitare che meriti il titolo di pianeta a tutti gli effetti. Stranamente, una serie di ricerche iniziate nella speranza di trovare un decimo pianeta potrebbe, in un certo senso, averne ridotto il numero finale a otto. Questa beffa della sorte, insieme con le molte interessanti osservazioni che abbiamo compiuto di oggetti della fascia di Kuiper, ci ricorda che il sistema solare serba ancora innumerevoli sorprese.

JANE X. LUU e DAVID C. JEWITT collaborano nello studio del sistema solare dal 1986, quando entrambi si trovavano al Massachusetts Institute of Technology. Oggi Jewitt lavora all'Università delle Hawaii e Luu alla Harvard University.

DUNCAN MARTIN, QUINN THOMAS e TREMAINE SCOTT, *The Origin of Short Period Comets* in «Astrophysical Journal», 328, pp. L69-L73, 15 maggio 1988.

LUU J. X., *The Kuiper Belt Objects* in *Asteroids, Comets, Meteors 1993*, a cura di A. Milani, M. Di Martino e A. Cellino, Kluwer Academic Publishers, 1993.

JEWITT D. C. e LUU J. X., *The Solar System beyond Neptune* in «Astronomical Journal», 109, pp. 1867-1876, aprile 1995.

MALHOTRA RENU, *The Origin of Pluto's Orbit: Implications for the Solar System beyond Neptune* in «Astronomical Journal», 110, pp. 420-429, luglio 1995.

La videochirurgia

Le tecniche di chirurgia mininvasiva si servono di strumenti miniaturizzati con i quali si praticano interventi che prima richiedevano ampie incisioni, dolorose e difficili da cicatrizzare

di François Dubois

Anestetizzare il paziente, penetrare nel suo organismo, eseguire l'intervento e rianimarlo: tutte queste diverse fasi di un'operazione chirurgica sono state rese meno traumatiche nel corso degli ultimi 20 anni. I progressi realizzati nei settori dell'anestesia e della rianimazione hanno permesso interventi nuovi; nel contempo la cosiddetta chirurgia videoendoscopica ha sconvolto la pratica chirurgica negli ultimi 15 anni. Questa tecnica è nata dalla volontà dei chirurghi di limitare le conseguenze fisiologiche dell'intervento, di minimizzare i rischi legati a un'ampia incisione dei tessuti e di accelerare la cicatrizzazione. La chirurgia videoendoscopica può quindi a buona ragione essere considerata una tecnica mininvasiva.

Nel corso di un intervento tradizionale, a cielo aperto, il chirurgo raggiunge tramite incisione l'organo interessato, che può osservare direttamente, toccare, tagliare o suturare prima di richiudere la ferita. Nel corso di un intervento mininvasivo, il chirurgo non osserva direttamente l'organo, ma soltanto la sua immagine trasmessa da un dispositivo ottico su uno schermo televisivo; non impugna direttamente gli strumenti, ma li comanda con regolatori esterni.

Questo metodo, di cui esamineremo i principi, le applicazioni, i vantaggi e anche gli inconvenienti, ha suscitato una specie di nuova *querelle* fra Antichi e Moderni. I pionieri della chirurgia mininvasiva si sono scontrati con la resistenza di molti colleghi; all'inizio i loro risultati erano rifiutati dalle riviste scientifiche internazionali. Essi hanno dovuto lavorare sotto il peso della disapprovazione ufficiale e subire molti sarcasmi: li si trattava addirittura da ciarlatani! Poi, nel 1990, negli Stati Uniti cominciò a diffondersi interesse per questo metodo e la chirurgia mininvasiva divenne più comune: oggi è un metodo riconosciuto e insegnato nelle facoltà di medicina.

Cinque piccole aperture

La chirurgia videoendoscopica discende dalla laparoscopia, un metodo che permette di osservare l'interno della cavità addominale. Nel 1902 un medico tedesco, G. Kelling, introdusse per primo un cistoscopio nella cavità peritoneale di un cane; lo strumento usato era un trocar (un tubo cavo) di qualche millimetro di diametro, munito di una piccola lampada all'estremità che veniva introdotta nell'addome e di un oculare all'altra estremità, attraverso il quale il medico compiva l'osservazione.

Kelling aveva dapprima iniettato aria nella cavità peritoneale per distendere le membrane che, aderendo normalmente all'intestino, ne avrebbero impedito l'osservazione. Qualche anno più tardi, nel 1910, lo svedese H. C. Jacobaeus fu a quanto pare il primo a esplorare con questo strumento le cavità polmonare e addominale dell'uomo. Tuttavia questi metodi, dapprima strettamente riservati all'osservazione dell'interno del corpo umano, rimasero quasi sconosciuti fino agli anni cinquanta; essi si svilupparono allora come ausilio diagnostico, in ginecologia e in patologia dell'apparato digerente e respiratorio.

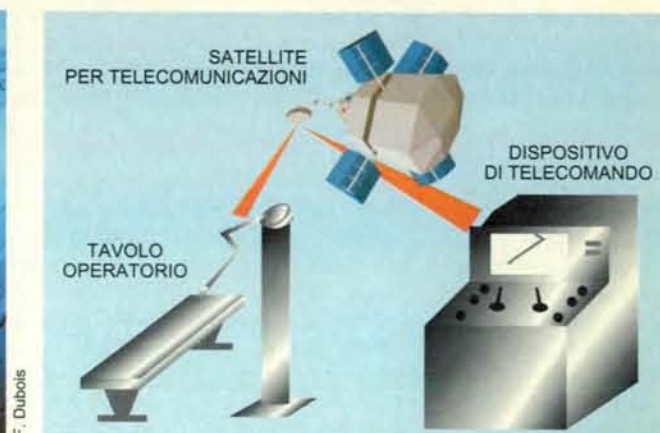
Ben presto alcuni interventi chirurgici vennero a completare la semplice osservazione: attraverso il trocar introdotto nella cavità addominale o toracica si osservavano gli organi e si cominciarono a sezionare tessuti fibrosi anomali della pleura, a prelevare campioni bioptici di tessuti, o ancora a perforare organi per asportarne, per esempio, una cisti.

Negli anni sessanta fu migliorato il sistema ottico: si sostituì la piccola lampada all'estremità del trocar con una sorgente di luce remota, trasmessa attraverso una fibra ottica; in questo modo i tessuti erano illuminati meglio e non venivano riscaldati dalla lampada. Negli anni settanta i ginecologi Kurt Semm di Kiel, Maurice Antoine Bruhat di Clermont-Ferrand e Hubert Manhès di Vichy eseguirono le prime operazioni chirurgiche su pazienti, intervenendo sulle ovaie e sulle tube di Falloppio. Alcune operazioni di chirurgia dell'apparato digerente furono praticate all'inizio degli anni ottanta, ma rimanevano ancora nel limbo della semi-ufficialità.

Nel 1987 Philippe Mouret, di Lione, forte della sua esperienza in laparoscopia, realizzò la prima ablazione celioscopica della cistifellea, o colecistectomia. Fin dai primi interventi, i chirurghi dovettero munirsi di tutta una gamma di strumenti di piccole dimensioni, che si potessero introdurre nell'addome attraverso i trocar. I gesti operatori (allontanare i tessuti, tagliare, suturare, far coagulare il sangue, arrestare il circolo, gli stessi che si compiono nella chirurgia tradizionale) richiedevano strumenti miniaturizzati: pinze, bisturi, porta aghi eccetera. Grazie a questa strumentazione si riuscirono a eseguire progressivamente con la chirurgia endoscopica le stesse operazioni della chirurgia classica.

In che cosa consiste questo metodo? Il chirurgo pratica quattro o cinque piccole incisioni (i punti di inserzione) per introdurre nell'addome del paziente anestetizzato trocar di diametro compreso fra cinque e dieci millimetri. Uno di essi contiene una fibra ottica collegata a una sorgente luminosa, la cui luce rende visibile il campo operatorio. Nel corso dei primi interventi il chirurgo osservava l'organo illuminato per mezzo di un visore telescopico contenuto nel trocar: teneva l'occhio fisso sull'oculare del visore, che reggeva con una mano, e operava con l'altra mano. Oggi il visore telescopico è collegato a una microcamera che riprende il campo; l'immagine trasmessa è proiettata su uno schermo televisivo, in modo che il chirurgo abbia entrambe le mani libere per lavorare, come in chirurgia classica. Generalmente l'ottica utilizzata permette l'osservazione lungo l'asse del visore, sul campo più ampio possibile; per accedere a certe zone si utilizzano visori che, grazie a una serie di specchi, consentono una visione laterale a 45 o anche a 90 gradi; endoscopi flessibili consentono l'esplorazione di ogni anfratto.

Operare su uno schermo televisivo

Il visore telescopico permette un'ampia esplorazione della cavità e l'osservazione degli altri strumenti. Oltre a questo trocar riservato all'illuminazione dell'organo bersaglio e alla sua osservazione, ve ne sono altri destinati agli strumenti: uno contiene i bisturi, un altro le pinze, un terzo un divaricatore. La videocamera riprende l'organo e gli strumenti in modo tale che il chirurgo

Durante un intervento di videochirurgia una microcamera è introdotta nell'addome del paziente e riprende l'organo bersaglio; l'immagine del «campo operatorio interno» viene osservata su uno schermo televisivo (*in alto a sinistra*). Il chirurgo manovra gli strumenti introdotti nell'addome mediante comandi situati sul trocar. Si può pensare che, tra qualche anno, il chirurgo non impugnerà più i comandi degli strumenti, ma li controllerà da una *consolle* dove saranno riuniti schermo di osservazione e comandi (*in basso*). Dato che il chirurgo non interviene direttamente sul paziente, salvo che per introdurre i trocar e preparare il campo operatorio interno, operazioni difficili, che richiedono specialisti a volte lontani, potrebbero essere eseguite con i mezzi di telecomunicazione più elaborati: l'immagine dell'organo operato, ripresa dalla videocamera, sarebbe inviata a un satellite intermedio che la ritrasmetterebbe al visore del chirurgo che opera (*in alto a destra*). Osservando il suo schermo, il chirurgo «interverrebbe» e i suoi ordini sarebbero a loro volta trasmessi dal satellite: che egli sia a un metro o a migliaia di chilometri di distanza non farebbe alcuna differenza. Questo almeno in teoria, dato che in pratica il chirurgo deve essere presente per ricorrere rapidamente alla chirurgia tradizionale in caso di necessità.

non osserva direttamente il campo operatorio, ma lo vede sullo schermo televisivo. Egli aziona gli strumenti per mezzo di comandi situati all'estremità dei trocar. I suoi strumenti sono quelli della chirurgia classica: bisturi, pinze, porta aghi, divaricatori, aspiratori, tutti adattati alla chirurgia endoscopica, ossia miniaturizzati, allungati, orientabili.

Rimane una difficoltà di cui non abbiamo ancora parlato: l'accesso all'organo e la sua visualizzazione. In chirurgia classica si apre la parete corporea e si crea così un campo operatorio a cielo aperto. In chirurgia mininvasiva questo campo operatorio è formato da una «bolla d'aria» creata insufflando gas nello spazio che separa i due foglietti peritoneali o pleurici. Si crea così uno pneumoperitoneo o uno pneumotorace, spazi vuoti artificiali dove gli strumenti possono muoversi liberamente.

Per insufflare il gas si utilizza l'ago di Veress, munito di un dispositivo di protezione che impedisce di lesionare gli organi una volta attraversata la membrana peritoneale (l'ago si copre automaticamente con un «cappuccio» protettivo). Il gas iniettato deve essere inerte e non tossico. Inoltre deve essere solubile nel sangue al fine di ridurre il rischio di embolia gassosa: se il gas non fosse solubile, le bolle trasportate dal sangue rischierebbero di ostruire vasi sanguigni polmonari o cerebrali. Di solito si usa anidride carbonica.

A causa di certi inconvenienti legati all'uso di questo gas, di cui parleremo in seguito, si sta tentando di mettere a punto sistemi di trazione meccanici che raggiungano lo stesso scopo. Tuttavia questi sistemi, nonostante i loro vantaggi teorici, sono ancora imperfetti, e non sollevano la parete altrettanto bene di un gas in pressione.

Nel corso di un intervento toracico l'uso del gas non è necessario: il polmone si affloscia spontaneamente allorché l'aria penetra nella cavità pleurica. Normalmente il polmone aderisce alla pleura: quando si inspira, questa - sospinta dalle coste - si solleva e i polmoni in depressione si riempiono d'aria. Quando l'aria entra nella pleura, l'azione meccanica normalmente esercitata dalle coste cessa e i polmoni non si riempiono più. Lo spazio fra il polmone afflosciato e le coste permette al chirurgo di operare con tecnica mininvasiva.

I campi di applicazione

Per due o tre anni, la chirurgia mininvasiva è stata quasi completamente riservata agli interventi sulla cistifellea. Poi, a partire dal 1989, ha esteso il suo raggio d'azione ad altri interventi sull'apparato digerente. Negli anni novanta la chirurgia videoendoscopica si è diffusa come un'onda di marea in tutto il mondo, e oggi viene applicata a quasi tutti gli interventi sull'apparato digerente, sui polmoni e sulla pleura, alla chirurgia ortopedica, alla neurochirurgia, alla chirurgia del naso, della gola e dell'orecchio, a quella dell'apparato urinario e perfino alla chirurgia plastica. Ci limiteremo qui a descrivere applicazioni di chirurgia dell'apparato digerente, polmonare e ginecologica.

L'assortimento degli strumenti usati in videochirurgia è simile a quella della chirurgia classica, ma tutti gli strumenti sono miniaturizzati. I trocar (*in alto al centro*) sono tubi metallici cavi attraverso i quali si introducono sia la sorgente luminosa che illumina il campo operatorio interno sia gli strumenti, come pinze e bisturi (*in basso*), con i quali il chirurgo opera. Questi strumenti sono lunghi da 40 a 50 centimetri.

In questi campi pressoché tutti gli interventi possono essere realizzati con tecnica mininvasiva, ma bisogna distinguere fra gli interventi che sono ormai appannaggio quasi esclusivo della chirurgia videoendoscopica e quelli che, se realizzati con questa tecnica, rappresentano vere e proprie imprese. Si praticano comunemente per via endoscopica operazioni sulle tube e sulle ovaie (quando non si devono asportare masse tumorali), l'ablazione della cistifellea, l'appendicectomia, la riduzione di ernie inguinali e, in chirurgia toracica, il trattamento di pneumopatie bollose o la pleurectomia (ablazione della pleura).

Viceversa, un chirurgo deve già avere una buona esperienza di chirurgia videoendoscopica per realizzare un'isterectomia (ablazione dell'utero) con questa tecnica, e altrettanto dicasi per il trattamento di ernie iatali (una protrusione dello stomaco nella cavità toracica con comparsa di maree acide), la splenectomia (ablazione della milza) o qualsiasi intervento di chirurgia pediatrica.

Infine, solo i centri molto specializzati praticano exeresi pancreatiche (ablazione del pancreas), gastrectomie (ablazione dello stomaco), resezioni polmonari (ablazione parziale di un lobo polmonare) e certi interventi sull'apparato cardiovascolare. Il gruppo di Alain Carpentier dell'Ospedale Broussais ha realizzato per la prima volta, il 26 febbraio 1996, un intervento a cuore aperto su una paziente colpita da gravi lesioni della valvola situata tra l'atrio e il ventricolo sinistro (mitrale o bicuspide).

La sala operatoria

Un intervento con tecnica mininvasiva richiede la messa in opera di una grande quantità di apparecchi in una sala operatoria già ingombra del tavolo operatorio, dei proiettori, del generatore per il bisturi elettrico, delle apparecchiature per l'anestesia, la radiologia e l'ecografia: perciò la chirurgia videoendoscopica necessita di locali di grandi dimensioni, perché si deve anche poter ricorrere in ogni momento, e a volte con urgenza, alla chirurgia classica (quando si manifesta una complicazione imprevista il chirurgo passa ai metodi tradizionali e apre l'addome).

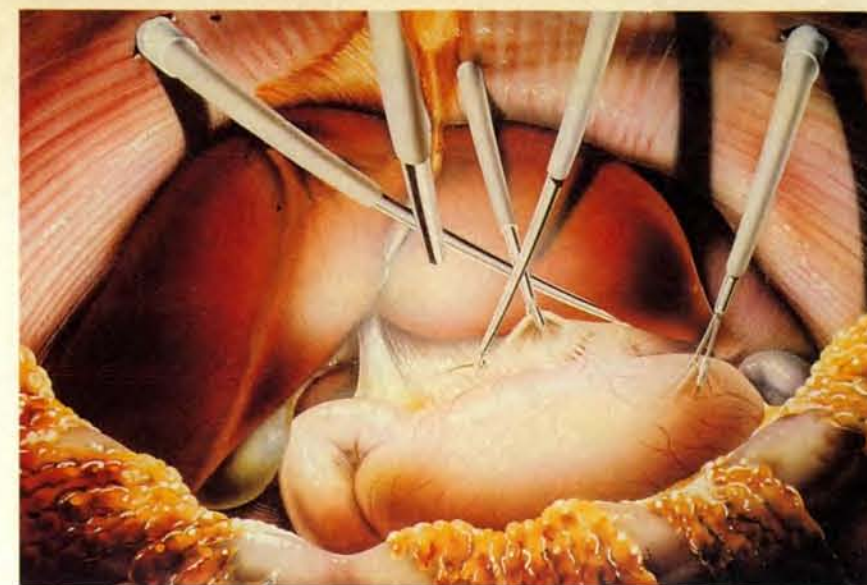
Gli strumenti specifici della videochirurgia sono l'insufflatore di gas, che insuffla anidride carbonica in quantità e a una pressione ben determinate e che

controlla automaticamente questi parametri durante tutto l'intervento; un dispositivo che inocula siero sotto pressione per ripulire la cavità dopo l'intervento; una sorgente di luce potente e di intensità regolabile; una videocamera e il suo monitor, che trasmette l'immagine sullo schermo, e un magnetoscopio che registra le immagini. La strumentazione comprende gli elementi necessari per creare e mantenere lo pneumoperitoneo nonché quelli per compiere l'intervento chirurgico vero e proprio.

In chirurgia si impiegano strumenti che sono ripuliti e sterilizzati dopo ciascun intervento e che hanno una durata di vita molto lunga, a volte superiore a quella del chirurgo stesso. Per la loro complessità e la loro fragilità, gli strumenti endoscopici sono difficili da pulire, sterilizzare e mantenere in perfetta efficienza. A ciò si aggiunge il timore della trasmissione di virus, come l'HIV o quello dell'epatite C, cosicché tende a diffondersi il materiale «usa e getta»: esso serve a una sola operazione, viene prodotto in serie, fornito già sterilizzato e pronto all'uso e poi gettato dopo l'intervento, cosa che risparmia fra l'altro molte ore di manodopera. Alcuni strumenti vengono forniti in kit monouso adatti a questo o quel tipo di intervento.

Generatori di ultrasuoni in miniatura possono essere introdotti attraverso i trocar e servire a scopi sia diagnostici sia di trattamento: essi permettono di realizzare ecografie *in situ*, o di frantumare con precisione calcoli biliari o urinari, dato che la sorgente di ultrasuoni viene portata a contatto del calcolo da distruggere. Il laser serve per tagliare, vaporizzare e coagulare. Il più utilizzato in ginecologia è il laser a CO_2, che permette una buona coagulazione superficiale, ma non viene trasmesso dalle fibre ottiche. Si può anche utilizzare un laser a impulsi per frantumare i calcoli e facilitarne l'estrazione. Tuttavia, a causa del suo costo e dell'esistenza di altre tecniche altrettanto efficaci, il laser tende a essere abbandonato per la chirurgia dell'apparato digerente.

Una chirurgia poco aggressiva

Le tecniche mininvasive, e soprattutto la chirurgia videoendoscopica, si sono particolarmente sviluppate negli ultimi anni perché presentano molteplici vantaggi. Il trauma legato all'intervento operatorio è ridotto rispetto a un'operazione eseguita con tecnica classica; la ferita operatoria è minuscola, e un'apertura di dieci centimetri, per esempio, è più dolorosa e più lunga da cicatrizzare che non cinque incisioni di due centimetri. Il dolore associato alle piccole cicatrici è moderato e non pregiudica il movimento e la respirazione; si

Veduta ideale di un campo operatorio interno: l'illustrazione rappresenta l'organo sul quale il chirurgo interviene (in questo caso lo stomaco) e gli strumenti che utilizza. La luce è trasmessa dal trocar situato in primo piano al centro. Gli altri strumenti sollevano i tessuti o li incidono. Questa immagine corrisponde a quella che viene trasmessa da una microcamera introdotta nell'addome per riprendere l'operazione.

evita il rischio di rottura della sutura e di riapertura della ferita; il trauma viscerale, peritoneale o pleurico è ridotto: i visceri restano nel loro ambiente naturale e non vengono raffreddati, né disseccati, né irritati dal contatto con compresse di garza; la loro manipolazione è minima. Grazie a questo fatto, le contrazioni intestinali sono poco perturbate e il transito riprende rapidamente.

L'operato avverte in scarsa misura dolori parietali, rigonfiamento postoperatorio e difficoltà di respirazione; può quindi alzarsi e alimentarsi rapidamente dopo la rianimazione, cosa che evita le complicazioni legate a una lunga permanenza a letto: cattiva circolazione sanguigna negli arti inferiori con rischio di flebite, rischio di infezioni polmonari o urinarie, che sono uno dei principali inconvenienti della chirurgia classica, soprattutto nei soggetti «a rischio» come anziani, portatori di handicap, cardiopatici e obesi.

Il periodo di dieta preoperatoria è ridotto e le abitudini alimentari dei pazienti devono essere modificate solo in maniera marginale. L'immobilità conseguente all'intervento operatorio è limitata: i periodi di ospedalizzazione e di assenza dal lavoro si riducono, il che costituisce un vantaggio essenziale del metodo, sia per il paziente sia per la società.

Nel corso dell'intervento l'osservazione e l'esplorazione sono di buona qualità, perché il dispositivo ottico utilizzato ingrandisce da cinque a quindici volte i tessuti osservati e rivela dettagli che l'occhio nudo non percepirebbe. Per di più, dato che il visore telescopico è orientabile in tutte le direzioni, esso trasmette le immagini di ogni anfratto, e il chirurgo che vede la totalità della cavità peritoneale può esplorarla altrettanto bene, se non meglio, che attraverso un'apertura classica anche molto grande.

Inoltre, a causa della piccolezza delle incisioni, si minimizzano i rischi di contaminazione dall'aria: nella chirurgia tradizionale la ferita rischia di essere infettata da microrganismi veicolati dall'aria o dal contenuto dei visceri. Le conseguenze dal punto di vista estetico sono minime: le cicatrici nei punti di incisione sono poco evidenti e vengono dissimulate in pieghe della pelle. Infine questo metodo riduce i rischi di aderenze postoperatorie: dopo un intervento la cicatrice interna crea tessuto fibroso che impedisce ai tessuti di scorrere liberamente nella cavità addominale; l'intestino, per esempio, rischia di arrotolarsi intorno a questi lembi di tessuto fibroso, con eventuale insorgenza di un'occlusione intestinale. Dato che il trauma è minore nel corso di un'operazione mininvasiva, questo tessuto fibroso è meno abbondante e i rischi di aderenze sono ridotti.

Gli inconvenienti

Il metodo, però, non è privo di potenziali inconvenienti. Per il chirurgo si tratta di una tecnica faticosa, che necessita estrema attenzione in ogni istante. Dopo alcune ore, teoricamente quattro, l'operatore dovrebbe essere sostituito;

sarebbero dunque necessari «cochirurghi», con lo stesso ruolo dei copiloti su un aereo. In più il chirurgo non osserva gli organi in rilievo, in tre dimensioni, ma su uno schermo bidimensionale; deve ricostruire mentalmente il rilievo, il che necessita di una certa abitudine. Non può aiutarsi con il senso del tatto, che nella chirurgia tradizionale è una buona fonte di informazioni, e l'assenza della palpazione manuale è mal compensata dagli strumenti. Un chirurgo abituato ai metodi classici deve «riciclarsi», e ciò risulta spesso difficile e oneroso.

Per il paziente l'intervento non è privo di qualche rischio, come lesioni di organi o vasi al momento dell'introduzione dei trocar, ma gli incidenti sono rari. Inoltre l'impiego di anidride carbonica in pressione - anche se la pressione è modesta e controllata di continuo - rischia di provocare un aumento della concentrazione di questo gas nel sangue, che si traduce in un innalzamento dell'acidità; nel corso di interventi lunghi, e in soggetti debilitati, l'iperacidità del sangue può indurre episodi di tachicardia. Infine, se il gas disciolto viene liberato bruscamente, esiste un rischio, sia pur minimo, di embolia gassosa.

Oltre alle complicazioni che possono manifestarsi nel corso dell'intervento, ve ne sono alcune che appaiono in seguito: si constata un aumento di complicanze locali, per esempio perdite biliari dopo una colecistectomia o fistole digestive (fuoriuscita di succhi digestivi in seguito a perforazioni della parete intestinale). In genere si ignora l'origine di queste complicanze, ma esse sono prodotte sovente dall'impiego di uno strumento non perfettamente adatto, di una tecnica approssimativa o anche - paradossalmente - da uno dei vantaggi sottolineati in precedenza: l'assenza di aderenze. Il tessuto fibroso che si forma dopo gli interventi di chirurgia tradizionale, se è a volte responsabile di occlusioni, serve da «colla» che riempie le eventuali piccole lacune della sutura. In assenza di questo tessuto fibroso, rischiano di apparire delle fistole. L'impiego della chirurgia videoendoscopica è controverso in casi di tumore maligno perché si pensa - senza averne la prova - che vi sia un rischio di disseminazione di cellule neoplastiche.

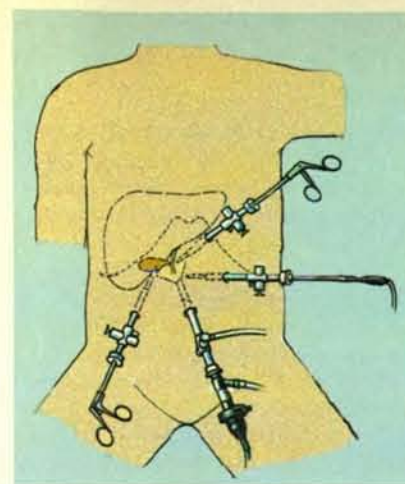
I costi della chirurgia videoendoscopica non sono ancora stimati con precisione, perché intervengono molteplici fattori di natura contrastante. Tuttavia la riduzione dell'ospedalizzazione e dell'assenza dal lavoro diminuisce indubbiamente i costi: un intervento praticato con videochirurgia può costare fino a due volte meno della stessa operazione eseguita con metodi classici.

Viceversa, l'installazione è onerosa: gli strumenti e l'equipaggiamento necessario (*si veda l'illustrazione in alto a destra a pagina 55*), costano circa 75 milioni di lire ed è prudente disporre di parecchie unità per far fronte a eventuali guasti. Il costo di mantenimento è elevato, ed è necessario rivolgersi a tecnici specializzati. Inoltre certi gruppi scelgono strumenti monouso, ma quale delle due tecniche (strumenti sterilizzati reimpiegabili o strumenti monouso) sia economicamente più vantaggiosa non è ancora stato stabilito.

Alcuni interventi durano più a lungo, impegnando maggiormente il personale e le sale operatorie. Un'ora di sala operatoria costa l'equivalente di due giorni di ricovero. L'evoluzione delle tecniche chirurgiche porterà cambiamenti anche nei centri ospedalieri: gli ospedali del futuro avranno meno posti letto, ma necessiteranno di basi tecnologiche sempre più complesse e costose.

I gesti del chirurgo

Come si diventa videochirurgo? L'insegnamento della chirurgia comporta un aspetto teorico, fornito dai corsi universitari e dai testi, e un aspetto tecnico artigianale, che si impara per apprendistato: il futuro chirurgo comincia osservando colui che opera, quindi lo aiuta, dapprima come secondo e poi come primo assistente; in seguito si fa assistere da un chirurgo esperto e infine opera da solo, chiedendo consiglio in caso di difficoltà. Questo periodo di apprendistato dura da quattro a cinque anni. Le cose vanno allo stesso modo anche per praticare la chirurgia videoendoscopica: ogni chirurgo deve ormai essere in grado di operare con le due tecniche e di passare dall'una all'altra.

In chirurgia videoendoscopica la gestualità non è naturale: il chirurgo non vede direttamente né gli organi su cui interviene né gli strumenti che utilizza. Segue i propri gesti su uno schermo bidimensionale e l'asse visuale è diverso dall'asse dei movimenti. Per giunta, i gesti sono capovolti: dato che lo strumento ruota intorno al punto di introduzione del trocar nei tessuti, si deve spingere verso sinistra quando si vuole che il bisturi vada verso destra. Per imparare si utilizzano manichini, e ci si allena a compiere i gesti della dissezione, della sezione, della legatura, delle suture, dei nodi eccetera. Tutto ciò è faticoso, per non dire umiliante, per un chirurgo abituato a fare questi gesti istintivamente ed elegantemente, e che torna a essere maldestro e pasticcione!

Come i piloti di linea si addestrano sui simulatori di volo, i chirurghi non potrebbero anch'essi esercitarsi su simulatori? No, almeno per ora, sia a causa del costo e del ristretto mercato di tali simulatori, sia in ragione di una difficoltà tecnica difficilmente superabile: contrariamente a una collina o a una pista di atterraggio fisse, gli organi addominali sono mobili e si modificano per effetto di ciascun gesto chirurgico. Realizzare modelli è difficile perché non si possono prevedere le reazioni e i movimenti degli organi. Viceversa, per organi immobili come il cervello, l'occhio, l'orecchio e il sistema osteoarticolare si può pensare a simulatori e a una chirurgia assistita dal calcolatore.

Se l'apprendimento dei gesti si fa su manichini, alcuni movimenti, quali la creazione di uno pneumoperitoneo o il controllo di una emorragia, devono essere realizzati su soggetti viventi; nonostante le questioni etiche sollevate dalla sperimentazione su animali, sembra preferibile che il chirurgo si addestri su un animale che su una persona.

La sperimentazione sugli animali resta indispensabile nella ricerca, nella messa a punto e nello studio delle conseguenze di queste tecniche di punta; i dati devono essere raccolti in mammiferi il più possibile vicini all'uomo sul piano fisiologico. Gli animali più utilizzati sono maiali di circa 30 chilogrammi; del resto, chirurghi di paesi come la Gran Bretagna, dove la sperimentazione su animali è vietata, vanno ad apprendere le tecniche di videochirurgia in Francia o in Germania.

La videochirurgia estenderà il proprio campo d'azione seguendo i progressi

Ph. Mouret

Per sollevare la parete addominale dagli organi da operare si può insufflare gas sotto pressione, in genere anidride carbonica. Tuttavia l'uso del gas non è privo di rischi, e alcuni preferiscono staccare meccanicamente la parete con un sospensore (*il principio è illustrato qui sopra*). Questo è fissato a un uncino la cui altezza regola quella della parete addominale (*a destra*).

Christian Klaiber

L'exeresi della cistifellea, o colecistectomia, comprende parecchie fasi: una volta introdotti nell'addome i diversi trocar (*in alto a sinistra*), la prima fase consiste nell'isolare la colecisti per mezzo di pinze (*a*) e nel liberare il dotto cistico (da *b* a *e*). Quando il dotto è liberato (*f*), lo si chiude per mezzo di graffe e si effettua il taglio fra le graffe (*g* e *h*). Bisogna poi fare un nodo intorno al moncone che rimane *in situ* (*i*). Si interrompe il flusso sanguigno e si taglia l'arteria cistica che irrora la cistifellea (*j*). Quest'ultima è ormai indipendente dal resto dell'organismo. Il chirurgo stacca completamente la ghiandola (*k* e *l*) e la estrae con uno dei trocar (*m*).

Si allarga il campo della chirurgia mininvasiva

Se si pensa che la prima asportazione per via laparoscopica della cistifellea risale a meno anni fa e che ora questo tipo d'intervento è eseguito di routine nella maggior parte degli ospedali, si può capire con quale rapidità riescano a diffondersi i progressi della tecnica quando conciliano più interessi. Paradossalmente, a fronte degli enormi vantaggi per il paziente (meno dolore postoperatorio, recupero più rapido e cicatrici minuscole), per il sistema sanitario (assistenza meno impegnativa, degenze più brevi e minori complicazioni) e per la società (meno assenze dal lavoro per interventi e convalescenza), finora era il chirurgo a dover sopportare il peso maggiore di una tecnica che, per l'innaturalità dei movimenti e la mancanza di contatto diretto con i tessuti del paziente, richiedeva un faticoso addestramento non solo manuale, ma anche e soprattutto mentale.

Ma tale impegnativa fase della chirurgia mininvasiva - ben descritta dal chirurgo francese autore di questo articolo - sta volgendo al termine. Ci si aspetta molto dal prossimo Convegno internazionale di chirurgia mininvasiva che si svolgerà a Cernobbio, tra il 18 e il 21 settembre prossimi e che vedrà riuniti i maggiori esperti mondiali sotto la presidenza di Enrico Pisani, direttore dell'Istituto di urologia del Policlinico di Milano, uno degli istituti all'avanguardia nell'applicazione delle più avanzate tecniche di intervento. Molte delle relazioni riguarderanno infatti gli studi e le ricerche cliniche che si prefiggono di ridare al chirurgo «mininvasivo» le percezioni che la chirurgia classica forniva all'operatore. Fred A. Voorhorst dell'Università di Delft (Paesi Bassi) porterà un contributo su alcuni accorgimenti tecnologici che, consentendo al chirurgo di collegare i movimenti del capo con l'obiettivo del laparoscopio e con i faretti che illuminano i tessuti esplorati, ricreano il contatto diretto tra chirurgo e campo operatorio. Ernest Flemming, dell'Università di Tubinga, descriverà un sensore vibrotattile che, mediante risonanza elettromagnetica, consente di distinguere la consistenza dei vari tessuti. Lo strumento è stato già messo alla prova con risultati positivi per individuare alcuni tipi di tumori come le metastasi epatiche degli adenomi del colon. Per ovviare a un inconveniente ricordato nell'articolo, la lesione incidentale di arterie, il tedesco Jörg-Uwe Meyer, di Ingbert, ha messo a punto un sensore pulsatile che rileva anche minuscole arterie non evidenti all'endoscopio e un sensore che imita il senso tattile posseduto dalla mano del chirurgo. Infine, per eliminare la fastidiosa mancanza di tridimensionalità dei sistemi di visione attuali, si stanno realizzando sofisticate apparecchiature collegate ai movimenti degli occhi del chirurgo il quale però, avverte Cletus von Pichler del Politecnico di Aachen, potrebbe trasformarsi in uno stressato utente di realtà virtuale.

Al Politecnico di Milano, la mano robotica si «allena» ad assistere il chirurgo.

E, sempre in clima di una fantascienza sempre meno fantastica, Peter Geis, del St. Joseph Hospital di Baltimora, parlerà del braccio robotico comandato dalla voce del chirurgo. Da bravo «ferrista», esso posiziona la strumentazione, illumina il campo, manovra le telecamere con la massima affidabilità ed evitando movimenti inutili.

Fin qui abbiamo citato studiosi stranieri, ma nel campo della robotica e della conseguente possibilità di eseguire interventi a distanza anche il nostro paese si trova in buona posizione: a Milano la collaborazione tra il Policlinico e Alberto Rovetta del Politecnico sta fornendo risultati concreti (ne abbiamo parlato sul fascicolo dello scorso novembre), e a Pisa il gruppo di P. Dario sta ottenendo importanti realizzazioni.

Un'altra importante applicazione delle tecniche mininvasive, sperimentata anche in Italia, riguarda la terapia: si è constatato che, somministrando antibiotici o antitumorali nel punto esatto in cui devono agire, si possono impiegare dosaggi molto più bassi e ottenere risultati migliori con minori effetti collaterali.

Il problema dell'addestramento

Se in futuro ci si avvarrà sempre più dei metodi mininvasivi per una gamma sempre più ampia di interventi e si potrà contare su una nuova generazione di chirurghi formatisi con queste tecniche, resta il problema della preparazione dei chirurghi attuali, che è stato affrontato al Congresso internazionale di endoscopia ginecologica (il primo della storia di questa disciplina) tenutosi a Parma alla fine di maggio.

«Oggi assistiamo a una generale impreparazione dei medici verso questo tipo d'intervento, tanto che solo alcuni grandi centri possono vantare una collaudata competenza. Non è raro infatti incontrare chirurghi che in sei mesi si improvvisano bravi laparoscopisti, ignorando i propri limiti e sottovalutando la difficoltà della tecnica. I ginecologi, dal canto loro, sebbene più cauti nell'avvicinarla, fanno ancora fatica a entrare nell'ottica della procedura per la loro differente formazione specialistica» ha sostenuto Bruno Andrei, della Divisione di ginecologia e ostetricia dell'Ospedale civile di Fidenza e responsabile nazionale dell'Associazione ostetrici ginecologi ospedalieri italiani (AOGOI) per la chirurgia endoscopica. Le difficoltà nascono proprio dalla diversa impostazione che bisogna acquisire per riuscire a operare con la nuova metodica.

Se tra dieci anni si potrà disporre di operatori più esperti, attualmente si paga ancora lo scotto della breve storia della laparoscopia con un gran numero di «neopatentati». Chi pratica questa tecnica non può infatti dimostrare la sua preparazione con l'esibizione di un attestato, perché non c'è una regolamentazione in materia. La laparoscopia oggi s'impara eseguendola, magari stando a fianco di un collega che ha già dimestichezza con la tecnica o facendo tesoro dell'infarinatura ricevuta a un corso universitario facoltativo frequentato durante la specialità. Su questo sfondo acquista un risalto maggiore l'iniziativa dell'AOGOI e della Società italiana di endoscopia e laserterapia ginecologica (SIELG) di creare le competenze necessarie organizzando validi corsi di specializzazione post-laurea, affinché la laparoscopia possa diventare una tecnica sempre più sicura e praticata da un numero crescente di medici. *(Adriana Giannini e Manuela Campanelli)*

tecnologici. Gli interventi che oggi sono vere e proprie imprese diventeranno comuni e sicuri. Alcuni inconvenienti attuali, per esempio l'impiego di anidride carbonica, saranno meglio gestiti, o anche eliminati. Il lavoro del chirurgo sarà facilitato dall'uso di strumenti più agevoli da maneggiare. Procedimenti semplici permetteranno di vedere in tre dimensioni: oggi è ancora necessario portare scomodi occhiali o fare ricorso, con scarso successo, all'olografia.

La chirurgia senza chirurgo?

I futuri dispositivi ottici miglioreranno la risoluzione e ingrandiranno il campo di osservazione per trocar di piccolo diametro. Portacamera telecomandati reagiranno a rivelatori fissati sul capo dell'operatore o saranno comandati a voce: al chirurgo basterà girare la testa verso destra oppure ordinare «destra» perché lo strumento «obbedisca». La telemanipolazione è anche oggetto di un programma di ricerca europeo nel quadro del progetto Eureka; micromotori e microrivelatori dovrebbero permettere di automatizzare certi interventi.

Così si è già riusciti a inserire protesi nel bacino con molta precisione grazie al calcolatore: l'immagine tridimensionale dell'osso sul quale deve avere luogo l'intervento viene digitalizzata grazie a radiografie e a immagini ottenute con lo scanner; il calcolatore definisce il canale da perforare perché la protesi si adatti perfettamente; grazie a micromotori comandati dal calcolatore l'osso viene forato e la protesi messa in sede.

Rimane un passo da compiere: se l'uomo è assistito da macchine abbastanza abili da sostituirlo, se il chirurgo non osserva più l'organo su cui interviene, ma una immagine su uno schermo, non occorre che egli si trovi in sala operatoria, e neppure nei pressi: potrebbe essere a migliaia di chilometri di distanza. I sistemi di trasmissione delle immagini con satelliti potrebbero consentire di affrancarsi dalla presenza diretta dell'operatore. Tuttavia i tempi di trasmissione delle informazioni (un decimo di secondo perché l'ordine venga comunicato) rischiano di essere superiori al tollerabile nel caso, per esempio, si scateni una emorragia. Infine il chirurgo può decidere in ogni momento di ricorrere alla chirurgia tradizionale.

I progressi della robotica, dell'intelligenza artificiale e della telematica aiuteranno senza dubbio i chirurghi a realizzare con sicurezza interventi la cui semplice menzione oggi apparirebbe azzardata, ma la chirurgia sarà ancora per molto tempo praticata da esseri umani.

La «colonna» impiegata negli interventi di videochirurgia comprende una sorgente luminosa, un dispositivo di controllo della pressione del gas insufflato, uno di ricostruzione delle immagini registrate dal sensore di osservazione, uno schermo che proietta queste immagini e un videoregistratore che le registra.

Pier/Götz

Per apprendere le tecniche di videochirurgia si usano modelli. Un chirurgo abituato alla chirurgia classica deve riapprendere tutti i gesti e abituarsi alle difficoltà del metodo: non vede direttamente l'organo operato, ma una sua immagine bidimensionale proiettata su uno schermo; non può aiutarsi con il tatto, che in chirurgia classica gli dà preziose informazioni, e deve abituarsi a capovolgere i gesti: dato che i trocar sono mobili attorno a un punto fisso (quello di penetrazione), quando vuole spostare lo strumento verso sinistra deve muovere il comando verso destra.

FRANÇOIS DUBOIS, membro della Académie de chirurgie, lavora agli Hôpitaux de Paris.

TESTAS J. e DELAITRE B., *Chirurgie digestive par voie cœlioscopique*, éditions Maloine, 1991.

MOURET PH., *La cholécystectomie endoscopique a 4 ans*, «Lyon Chirurgical», 87, p. 179, 1991.

BRUHAT M. A. e DUBOIS F., *Cœlio-chirurgie abdomino-pelvienne*, Springer-Verlag, 1992.

CUSHIERI A., BUESS G. e PERISSAT J., *Operative Manual of Endoscopic Surgery*, Springer-Verlag, 1992.

La «Torre pendente» di Pisa

L'inclinazione del celebre campanile è andata continuamente aumentando dal XII secolo, ma con l'applicazione delle tecnologie più moderne oggi si spera di salvare questo magnifico monumento

di Paolo Heiniger

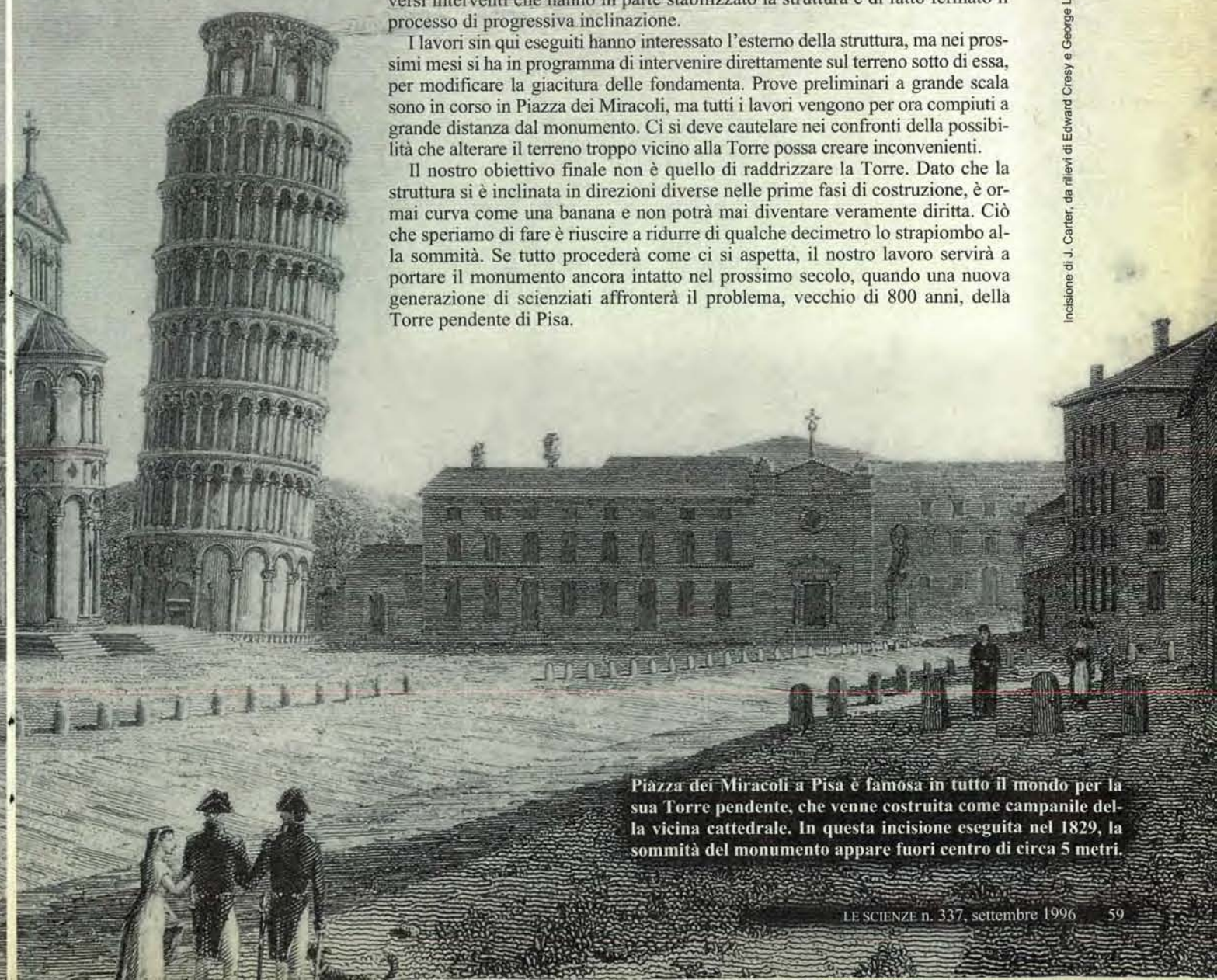
La «Torre pendente» non è mai stata realmente diritta: le fondamenta cedettero poco dopo l'inizio dei lavori nel 1173 e l'edificio cominciò a inclinarsi verso nord. Quando la costruzione venne proseguita, dopo un'interruzione di quasi un secolo, le fondamenta diedero nuovamente segni di movimento, sicché nel 1272 la Torre pendeva visibilmente verso sud. Oggi la sua sommità è 5,227 metri fuori centro e tuttora inclinata a mezzogiorno.

Per tutta la vita del monumento architetti e ingegneri hanno cercato di intervenire in vari modi, ma dal 1911, cioè da quando sono iniziati controlli regolari di cui rimangono documenti, la pendenza è aumentata con un gradiente abbastanza costante, di circa 1,2 millimetri all'anno. I timori per la sorte dell'edificio si acuirono quando la Torre civica di Pavia, di analoghe dimensioni, crollò all'improvviso nel 1989. Poco dopo la Torre di Pisa venne chiusa ai visitatori.

Nel 1990 il Governo italiano nominò una speciale commissione, il Comitato di consulenza per la salvaguardia della Torre di Pisa, composta di esperti italiani e stranieri nei campi dell'ingegneria strutturale, dell'ingegneria geotecnica, della storia dell'arte e del restauro dei monumenti, allo scopo di studiare nuovi metodi per salvare la Torre. Il gruppo nel quale lavoro, il Consorzio Progetto Torre di Pisa, seguendo le direttive del Comitato, ha messo in atto diversi interventi che hanno in parte stabilizzato la struttura e di fatto fermato il processo di progressiva inclinazione.

I lavori sin qui eseguiti hanno interessato l'esterno della struttura, ma nei prossimi mesi si ha in programma di intervenire direttamente sul terreno sotto di essa, per modificare la giacitura delle fondamenta. Prove preliminari a grande scala sono in corso in Piazza dei Miracoli, ma tutti i lavori vengono per ora compiuti a grande distanza dal monumento. Ci si deve cautelare nei confronti della possibilità che alterare il terreno troppo vicino alla Torre possa creare inconvenienti.

Il nostro obiettivo finale non è quello di raddrizzare la Torre. Dato che la struttura si è inclinata in direzioni diverse nelle prime fasi di costruzione, è ormai curva come una banana e non potrà mai diventare veramente diritta. Ciò che speriamo di fare è riuscire a ridurre di qualche decimetro lo strapiombo alla sommità. Se tutto procederà come ci si aspetta, il nostro lavoro servirà a portare il monumento ancora intatto nel prossimo secolo, quando una nuova generazione di scienziati affronterà il problema, vecchio di 800 anni, della Torre pendente di Pisa.

Incisione di J. Carter, da rilievi di Edward Cresy e George L. Taylor

Piazza dei Miracoli a Pisa è famosa in tutto il mondo per la sua Torre pendente, che venne costruita come campanile della vicina cattedrale. In questa incisione eseguita nel 1829, la sommità del monumento appare fuori centro di circa 5 metri.

Le fondamenta

Gli strati di argilla e sabbia che costituiscono il suolo sotto la Torre si sono consolidati in modo diseguale per effetto del peso, causando l'inclinazione del monumento. I primi sette metri circa sotto la struttura sono costituiti da una mescolanza di limo, argilla e sabbia. Più in basso, fino a circa 20 metri di profondità, vi è uno strato argilloso chiamato localmente «pancone», notevole per il suo colore grigio-azzurro. Il suolo sabbioso che divide questi primi due strati è orizzontale sotto gran parte di Piazza dei Miracoli, ma mostra una depressione a coppa in corrispondenza della Torre. Strati compressibili, prevalentemente argillosi, si incontrano fino a 40 metri circa, dopodiché si trova sabbia compatta e, a profondità maggiori di 70 metri, altri strati di materiale più rigido.

LIMO, ARGILLA E SABBIA

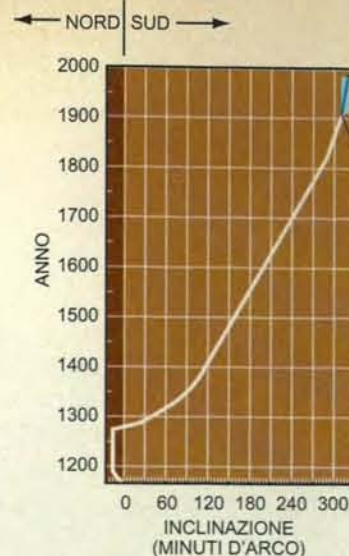
PANCONE (ARGILLA)

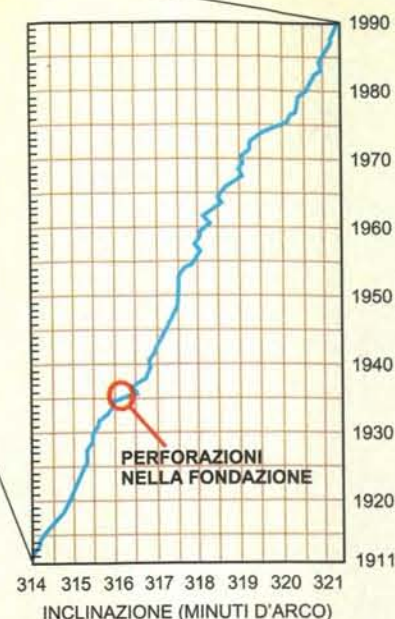
ARGILLA

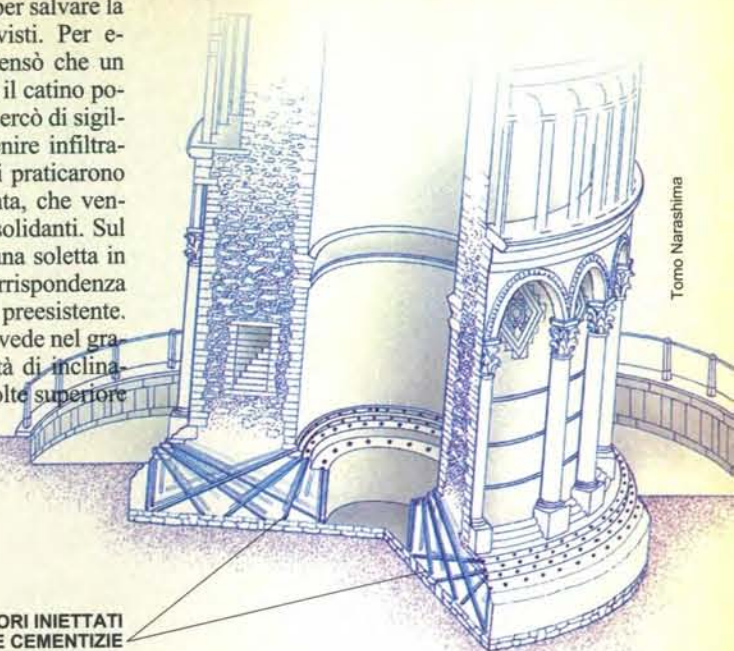
SABBIA

ARGILLA

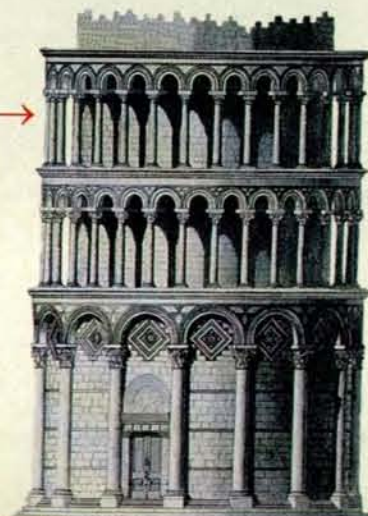
SABBIA LIMOSA

Tomo Narashima

Otto secoli di inclinazione

Inizialmente la Torre pendeva verso nord, ma per gran parte della sua storia è andata poi inclinandosi verso sud. La velocità di aumento della pendenza è stata massima nella prima parte del XIV secolo. Dal 1911, quando sono iniziati i controlli regolari, al 1990, la sommità della Torre si è spostata in media di circa 1,2 millimetri all'anno. Nel 1993 le misure prese recentemente per controllare l'aumento della pendenza hanno cominciato a produrre effetti.

Interventi alla base della Torre

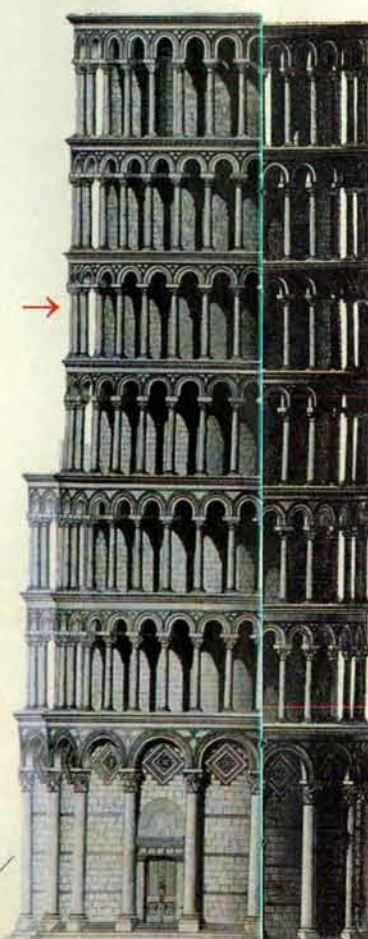
In passato alcuni dei tentativi fatti per salvare la Torre hanno avuto effetti imprevisti. Per esempio, dopo che, nel 1935, si pensò che un continuo flusso d'acqua attraverso il catino potesse indebolire le fondamenta, si cercò di sigillare la base della Torre per prevenire infiltrazioni dal basso. A questo scopo si praticarono obliquamente fori nelle fondamenta, che vennero poi iniettati con miscele consolidanti. Sul fondo del catino venne costruita una soletta in cemento armato, demolendo in corrispondenza del suo spessore il conglomerato preesistente. A seguito di questi lavori, come si vede nel grafico a sinistra, nel 1935 la velocità di inclinazione balzò a un valore oltre sei volte superiore a quello dell'anno precedente.

Tomo Narashima

La prima fase: 1173-1178

Il campanile del Duomo di Pisa si inclinò gradualmente durante la costruzione, avvenuta in tre fasi che coprirono un arco di tempo di 200 anni. I lavori vennero iniziati nel 1173, e già durante le prime fasi costruttive il monumento si piegò leggermente verso nord. Se ne ha una prova nel tentativo di correzione escogitato dagli architetti: per mantenere orizzontali i piani della Torre, gli archi e le colonne del terzo livello furono costruiti un po' più alti verso nord, dove il terreno stava cedendo, che non verso sud. A causa di rivolgimenti politici avvenuti a Pisa, la costruzione fu bloccata nel 1178, a metà dei lavori per il quarto ordine di colonne.

INCLINAZIONE DELLA TORRE NEL 1178

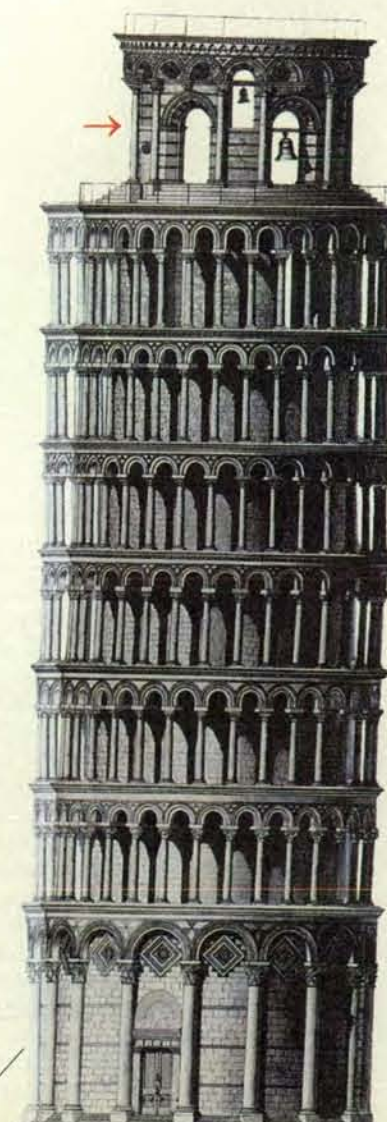
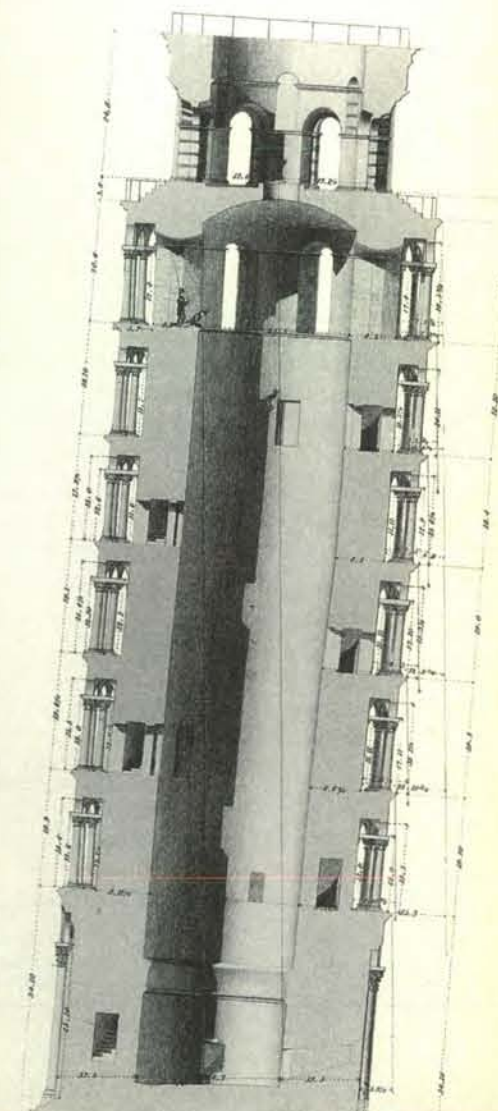
La seconda fase: 1272-1278

Alla ripresa dei lavori dopo quasi un secolo, nel 1272, l'inclinazione della Torre si era spostata verso sud, la direzione nella quale pende ancora oggi. Gli architetti tentarono una nuova correzione, questa volta modificando l'altezza del quinto livello e realizzando archi e colonne del lato sud un po' più alti di quelli del lato nord. Nel 1278, dopo che erano stati completati sette piani, i lavori furono sospesi nuovamente a causa della situazione politica. Nel 1292 l'inclinazione della Torre era ormai così pronunciata che un gruppo di mastri muratori fu incaricato di studiare il problema: la prima delle molte commissioni che si sono occupate del monumento nel corso di 700 anni.

INCLINAZIONE NEL 1178

INCLINAZIONE NEL 1278

La terza fase: 1360-1370

L'ottavo piano e la cella campanaria furono costruiti fra il 1360 e il 1370. Ancora una volta gli architetti cercarono di correggere la pendenza, questa volta inclinando verso nord la cella campanaria. Le correzioni effettuate durante le varie fasi costruttive si individuano meglio esaminando la Torre in sezione. Questi espedienti, combinati con il lungo periodo di costruzione (che ha dato al terreno sottostante la fondazione il tempo di assestarsi e quindi di rafforzarsi a sufficienza per compensare l'inclinazione), hanno finora impedito il crollo del monumento.

INCLINAZIONE NEL 1278

INCLINAZIONE NEL 1370

INCLINAZIONE ATTUALE DELLA TORRE

Incisione di J. Carter, da rilievi di Edward Cresy e George L. Taylor

Tentativi moderni di rafforzare la Torre

Il Comitato internazionale incaricato di salvare il monumento ha studiato e definito vari interventi di stabilizzazione. Per esempio l'entità dell'inclinazione impone una intensa sollecitazione sulla muratura del lato meridionale, che potrebbe provocare prima o poi il crollo dell'edificio. Nel 1992 il secondo piano della costruzione è stato rinforzato con una cerchiatura costituita da una serie di cavi d'acciaio per prevenire un ulteriore indebolimento della muratura in pietra.
Essendo stata presa questa misura precauzionale, il Comitato sta ora studiando come fermare definitivamente l'inclinazione della Torre. Sul lato settentrionale di un primo anello di calcestruzzo armato precompresso, costruito a cingere la base dell'edificio, sono state posizionate oltre 750 tonnellate di lingotti di piombo. Questi contrappesi hanno arrestato l'inclinazione della Torre e spostato verso nord la sua sommità di circa 2,5 centimetri in un periodo di nove mesi, da giugno 1993 a febbraio 1994.

CERCHIATURA CON CAVI D'ACCIAIO

CONTRAPPESI DI PIOMBO

SISTEMA DI PERFORAZIONE PER L'ESTRAZIONE DI TERRENO (PROPOSTO)

ANELLO DI CALCESTRUZZO ARMATO (IN COSTRUZIONE)

ELETTRODI INFISSI NEL TERRENO PER L'ESTRAZIONE DI ACQUA DALL'ARGILLA TRAMITE ELETTROSMOSI (PROPOSTO)

TIRANTI D'ACCIAIO CHE ANCORANO L'ANELLO DI CALCESTRUZZO ARMATO A UNO STRATO PROFONDO DI TERRENO SABBIOSO (PROPOSTO)

Tomo Narashima

Nel giugno 1995 i tecnici hanno cominciato a installare un secondo anello di calcestruzzo armato precompresso attorno alla Torre. Esso verrà ancorato a uno strato di terreno sabbioso a 50 metri di profondità per mezzo di cavi di acciaio sul lato settentrionale. Questo secondo anello con gli ancoraggi è destinato a sostituire i contrappesi di piombo, e dovrebbe esercitare un effetto stabilizzante ancora maggiore. Si prevede che in questo modo la sommità del monumento si sposterà di almeno altri 2,5 centimetri.
Si stanno prendendo in considerazione tecniche più aggressive, che altererebbero le caratteristiche del suolo sotto la Torre. Per esempio si è studiata la possibilità di utilizzare l'elettrosmosi per consolidare gli strati di argilla fra 10 e 20 metri di profondità. Elettrodi di grandi dimensioni realizzati con tubi di acciaio infissi nel terreno genererebbero nello strato argilloso un campo elettrico capace di attrarre in posizioni prefissate acqua, che si potrebbe così estrarre dall'argilla sotto la parte settentrionale delle fondamenta. Gli strati argillosi si comprimerebbero, favorendo una rotazione verso nord della Torre.
In alternativa si potrebbero impiegare sofisticati metodi di perforazione per estrarre piccoli volumi di terreno da sotto il lato settentrionale, riducendo lo spessore degli strati sottostanti e inducendo la Torre ad assumere un assetto più stabile.

Philip Vanoutrive/SIPA

Oltre 750 tonnellate di piombo sono state collocate sul lato settentrionale della base della Torre. I contrappesi hanno fatto ruotare la fondazione verso nord, permettendo alla costruzione di spostarsi di un paio di centimetri.

Massimo Sestini/Gamma Liaison

L'interno della Torre ora ospita apparecchiature di controllo molto sensibili in grado di rilevare minime variazioni nell'inclinazione. Nel settembre 1995, per esempio, gli scienziati hanno osservato che la sommità dell'edificio si è spostata verso sud di 0,24 millimetri nell'arco di due giorni.

Gianni Giansanti/Sygma

Una cerchiatura costituita da cavi di acciaio cinge il secondo piano, che rischiava di cedere a causa della pressione esercitata su questa parte del monumento. Per effetto dell'inclinazione, gran parte del peso dell'edificio (14 700 tonnellate) grava sul lato meridionale.

PAOLO HEINIGER è direttore in cantiere del Consorzio Progetto Torre di Pisa e ha avuto la responsabilità dell'esecuzione dei lavori sul monumento negli ultimi tre anni. È laureato in ingegneria al Politecnico di Milano e ha partecipato in Italia, Estremo Oriente, Regno Unito e Scandinavia all'esecuzione di numerosi lavori nel campo dell'ingegneria geotecnica e delle fondazioni di strutture in generale.

Le incisioni che illustrano questo articolo sono tratte dal libro *Architecture of the Middle Ages in Italy: Illustrated by Views, Plans, Elevations, Sections, and Details, of the Cathedral, Baptistery, Leaning Tower or Campanile, and Camposanto at Pisa: From Drawings and Measurements Taken in the Year 1817* di Edward Cresy e George L. Taylor, pubblicato a Londra nel 1829.

Massimo Sestini/Gamma Liaison

I compact disc del futuro

Un accordo recentissimo tra società concorrenti nel settore dell'elettronica ha dato origine a un piano innovativo per il lancio di prodotti compatibili DVD, a partire da questo autunno

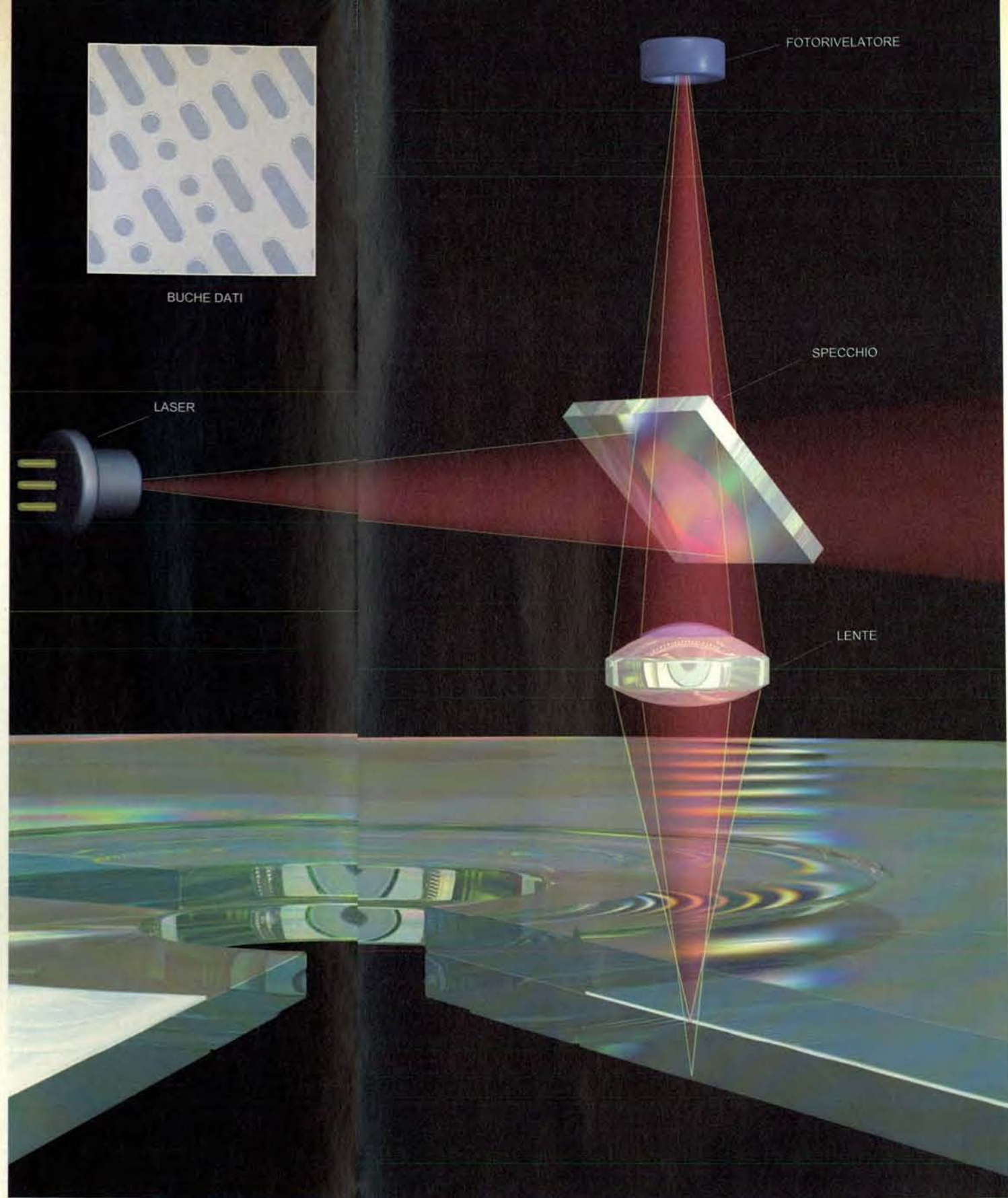
di Alan E. Bell

I compact disc e i loro lettori sono tra i prodotti di maggior successo nella storia dell'elettronica di consumo. Dal 1982, data di lancio dei prodotti con tecnologia CD audio, sono stati venduti più di 400 milioni di lettori e sei miliardi di dischi. Il CD-ROM (*read-only memory*, o «memoria a sola lettura»), un'estensione del formato CD audio originale, ha ottenuto un successo altrettanto notevole nelle applicazioni destinate ai personal computer: si prevede che nel solo 1996 si venderanno oltre 35 milioni di lettori. E presto vedrà la luce la seconda generazione di questa tecnologia, basata sul formato DVD (*digital versatile disc*, o «disco digitale versatile»). Dieci grandi società del settore elettronico hanno in programma di presentare una serie di prodotti a tecnologia DVD - compresi riproduttori video e lettori di DVD-ROM - negli ultimi mesi di quest'anno e nei primi del 1997.

Il formato DVD è il frutto di un accordo concluso verso la fine del 1995 tra gruppi concorrenti di società internazionali, che sono riusciti a combinare le caratteristiche migliori delle impostazioni elaborate indipendentemente da ciascuno. La nuova generazione di lettori prevista dall'accordo funzionerà sia con i CD esistenti sia con i DVD, i quali possono immagazzinare, grazie a tutta una serie di innovazioni progettuali, circa 14 volte più informazioni dei CD attuali. Inoltre la velocità con cui i lettori DVD di prima generazione restituiranno i dati - 11 milioni di bit al secondo - sarà pari a quella di uno dei più veloci lettori di CD-ROM attuali, e quindi rappresenterà uno standard di prestazioni completamente nuovo.

Come è facile immaginare, prestazioni e capacità così elevate si adattano a una gamma di applicazioni davvero notevole. I DVD possono registrare musica, film, giochi e altri prodotti multimediali, proprio come avviene già con i CD, ma saranno molto più capienti e la qualità di riproduzione sarà di gran lunga superiore. Inoltre dovrebbero dar modo di concepire prodotti completamente nuovi. Un video realizzato per un lettore di DVD, per esempio, potrebbe non limitarsi a contenere tutto un film, ma anche offrire allo spettatore diverse scelte di inquadratura, trama o lingua del doppiaggio. Si prevede un buon successo anche per i programmi di karaoke interattivo. Ed entro alcuni anni dovrebbero comparire sul mercato dischi e lettori di DVD-RAM (*random-access memory*, «memoria ad accesso diretto») e DVD-R (*recordable*, «registrabili»). Spingendosi ancora più in là con le previsioni, credo che dovrebbero entrare in commercio anche videocamere digitali basate su dischi DVD-RAM registrabili.

Il formato DVD e quello CD sono basati sulla stessa tecnologia di registrazione ottica: le informazioni sono rappresentate da «buche» microscopiche formate sulla superficie del disco di plastica quando il materiale è iniettato nello stampo. Il lato del disco con le buche viene poi ricoperto di uno strato sottile di alluminio seguito, nel caso di un CD, da uno strato di lacca protettiva e da un'etichetta. Per leggere i dati, il lettore proietta un piccolo fascio laser, attraverso il substrato del disco, sullo strato che contiene i dati, mentre il disco ruota su se stesso. L'intensità della luce riflessa dalla superficie varia a seconda che il raggio laser incontri o no una buca lungo la pista che contiene le informazioni. Quando c'è una buca direttamente sotto il fascio di lettura, il disco riflette molta meno luce di quando il fascio illumina un tratto piano della pista. Un fotorivelatore e altri dispositivi elettronici all'interno del lettore traducono la variazione di intensità luminosa negli 0 e negli 1 di un codice digitale che rappresenta le informazioni registrate sul disco.

Michael Goodman

Diversamente dai CD, i nuovi dischi a registrazione ottica contengono due strati di buche che rappresentano i dati (*ingrandite nel riquadro*). Mentre il disco ruota, questi minuscoli incavi passano sotto un fascio di luce laser (*in rosso*) e fanno variare la quantità di luce che il disco riflette. Un fotorivelatore e altri dispositivi elettronici traducono queste variazioni negli 0 e negli 1 che costituiscono i dati in forma digitale. Se si regola la posizione della lente, il dispositivo è in grado di leggere informazioni indifferentemente dallo strato superiore o da quello inferiore del DVD. La luce che attraversa l'ologramma al centro della lente si mette a fuoco su un secondo punto, adatto alla lettura dei CD attualmente in commercio.

Ci sono due differenze fisiche fondamentali tra CD e DVD. Innanzitutto le buche più piccole sui DVD hanno un diametro di soli 0,4 micrometri, mentre le buche corrispondenti di un CD sono grandi il doppio, cioè 0,83 micrometri. Inoltre le piste dati di un DVD distano solo 0,74 micrometri l'una dall'altra, mentre quelle dei CD si trovano a 1,6 micrometri di distanza. Per questo motivo, nonostante le dimensioni dei dischi siano identiche, la pista dati di un DVD è lunga più di 11 chilometri, ossia oltre il doppio di quella di un CD. Per poter leggere le buche più piccole, il fascio di lettura di un lettore di DVD deve essere messo a fuoco più finemente di quello di un lettore di CD. Per questo si usa un laser rosso a semiconduttore con una lunghezza d'onda compresa tra 635 e 650 nanometri, mentre i lettori di CD impiegano laser infrarossi con una lunghezza d'onda di 780 nanometri. Inoltre i lettori DVD utilizzano una lente più potente per la messa a fuoco. È a queste differenze, oltre che all'efficienza superiore (come vedremo più oltre), che il formato DVD deve la capacità di ben 4,7 gigabyte di ciascuno strato di informazioni.

Questa capacità può essere raddoppiata, fino a 9,4 gigabyte (e quasi quadruplicata, portandola a circa 17 gigabyte), grazie ad altre due innovazioni. I DVD e i CD hanno lo stesso spessore totale, pari a 1,2 millimetri, ma i DVD possiedono due substrati capaci di contenere informazioni, mentre i CD ne hanno uno solo. I substrati di un DVD sono uniti insieme in modo che le superfici con le buche siano rivolte l'una verso l'altra al centro del disco. Questa disposizione ripara le superfici dagli effetti deleteri di particelle di polvere e graffi. Nella versione più semplice, per accedere all'altra faccia di un DVD bisogna toglierlo dal lettore, capovolgerlo e reinserirlo, ma esiste una variante - chiamata multistrato - che permette di riprodurre tutte e due le superfici accedendo al disco dallo stesso lato.

In un disco multistrato il substrato superiore è rivestito di uno strato parzialmente riflettente e parzialmente trasmittente. La riflettività dello strato superiore è sufficiente per consentire al laser di leggere le buche nel substrato superiore, mentre la sua trasmissività permette di mettere a fuoco il fascio sul substrato inferiore e leggerne le buche. Quando il laser è a fuoco sulle buche dello strato inferiore, quelle dello strato superiore sono fuori fuoco e quindi non

danno fastidio. (Per tener conto della riduzione, piccola ma inevitabile, nella qualità di riproduzione, è necessario ridurre leggermente la capacità a 8,5 gigabyte. Questo spiega perché un DVD a doppia faccia e doppio strato contenga circa 17 gigabyte.) Per unire i due substrati occorre una colla speciale, e lo spessore dello strato legante va controllato con precisione per evitare eccessive aberrazioni nella messa a fuoco del fascio di lettura.

Il modello DVD a due substrati fornisce altri vantaggi, oltre alla maggiore capacità: riduce infatti gli errori causati dall'eventuale inclinazione e deformazione del disco. Tutti i compact disc sono soggetti a deformarsi, e quando la superficie di un disco si inclina e non giace più perpendicolarmente al fascio laser possono verificarsi errori di lettura. La gravità del degrado del fascio di lettura dovuto all'inclinazione è direttamente proporzionale allo spessore del substrato; quindi lo spessore di soli 0,6 millimetri del substrato dei DVD contribuisce a migliorare l'efficienza del sistema in generale. La sottigliezza del substrato rende i DVD meno sensibili all'inclinazione rispetto ai CD, che hanno un substrato di spessore doppio.

Esistono altre ragioni che rendono il DVD intrinsecamente meno vulnerabile a certi tipi di deformazione. Per esempio, variazioni improvvise della temperatura e dell'umidità possono dare luogo a rigonfiamenti o restringimenti dei substrati plastici di un DVD; ma, data la simmetria della sua struttura generale, le variazioni di uno degli strati tendono a compensare quelle dell'altro, riducendo l'effetto complessivo dei cambiamenti ambientali e minimizzando l'entità della deformazione che ne deriva.

Dato che i consumatori hanno già investito somme significative nelle loro collezioni di CD audio e di CD-ROM, si è data la massima priorità in fase di progettazione alla capacità dei lettori di DVD di leggere i CD esistenti oltre a quelli nuovi. A questo scopo è stato necessario introdurre alcune caratteristiche ottiche particolari nel progetto. La soluzione più semplice consiste nel montare due lenti su un'unica testina ottica - una lente ottimizzata per substrati di 1,2 millimetri e una ottimizzata per substrati di 0,6 millimetri di spessore - e poi passare meccanicamente da una lente all'altra quando è necessario.

Una soluzione più elegante è quella di usare una singola ottica stampata, avente un elemento olografico nel centro. La luce che attraversa la regione anulare esterna della lente non risente dell'ologramma e viene focalizzata in un fascio di dimensioni ridotte, adatto alla lettura dei DVD. Circa un terzo del fascio di lettura incidente sulla regione centrale viene messo a fuoco sia dalla lente sia dall'ologramma e forma un fascio adatto a leggere le buche dei CD.

Oltre ad avere più buche dei CD, i DVD concentrano più informazioni all'interno delle buche stesse grazie a miglioramenti in due aspetti dell'efficienza di codificazione del formato. Qualunque sia la forma originale delle informazioni (dati, testo, immagini, audio o video) gli 0 e gli 1 digitali che rappresentano il contenuto, i «bit utente», vanno protetti dagli effetti di errori introdotti durante la lettura. Questi errori derivano da cause quali polvere, graffi e corrosione. Le tecniche di controllo e correzione degli errori (ECC) minimizzano questi problemi grazie ad algoritmi speciali che calcolano bit di dati aggiuntivi da registrare in aggiunta ai dati utente. Questi bit addizionali, per quanto essenziali, riducono la percentuale della capacità totale disponibile per ospitare il contenuto vero e proprio.

Ma l'ECC dei DVD è molto potente. Per esempio, può correggere una serie di errori lunga fino a 2000 byte, pari a 4 millimetri di lunghezza della pista. Nel formato DVD i dati di ECC occupano circa il 13 per cento della capacità del disco, mentre nei CD essi richiedono ben un terzo della capacità. Il miglioramento di efficienza dell'ECC dei DVD, senza alcuna diminuzione delle capacità di correzione degli errori, è dovuto in gran parte alla maggiore potenza di calcolo dei circuiti integrati odierni, che non era disponibile all'epoca della progettazione del formato CD.

Durante la registrazione, la combinazione dei dati utente e di quelli di ECC va convertita nei cosiddetti bit di codice di modulazione, che costituiscono il flusso di bit effettivamente rappresentato dalle buche sulla superficie del disco. Questo passo è necessario per controllare l'intervallo di dimensioni delle buche occorrenti per rappresentare i dati, un fattore importante per garantire che durante la lettura i dati vengano rilevati e riprodotti senza errori. Il metodo di codifica del formato CD trasforma 8 bit utente in 17 bit di codice di modulazione, mentre il formato DVD usa un metodo che trasforma 8 bit utente in soli 16 bit di codice di modulazione, senza nel contempo perdere i benefici inerenti al metodo del formato CD. Dato che occorrono meno bit di codice di modulazione per rappresentare i bit dei dati utente, il DVD può contenere più dati utente; questa caratteristica porta a un miglioramento di efficienza di circa il 6 per cento rispetto al formato CD.

Le vendite di lettori di CD (compresi i CD-ROM e altri formati apparsi più di recente) continuano a salire costantemente dal 1982, anno in cui questi apparecchi furono introdotti sul mercato. Nel grafico, il numero di lettori venduti negli ultimi 14 anni.

Il formato DVD permetterà per la prima volta di immettere sul mercato lungometraggi - completi di colonna sonora con un livello qualitativo da sala cinematografica - pressoché indistinguibili dal nastro *master* originale. Questa possibilità solleva importanti questioni di protezione del diritto d'autore nel mondo digitale, un problema oggetto di continue discussioni. Resta però innegabile che capacità e prestazioni del DVD sono state scelte proprio in modo da fare della riproduzione di film uno dei suoi impieghi principali.

Nonostante i ben 4,7 gigabyte di capacità di ciascuna superficie di un DVD, è necessario usare un formato compresso per far stare un intero film digitale su un solo disco. A questo fine si usa un metodo di compressione a velocità del flusso di dati variabile, lo standard MPEG2, che permette di adattare in maniera ottimale il livello di compressione alla complessità della scena in ogni momento: le sequenze molto ricche di particolari o con azioni di movimento veloce tendono a essere più difficili da comprimere, e quindi si assegnano loro più bit che alle scene meno complesse. Come conseguenza, la velocità del flusso di dati compresso varia durante la riproduzione. La compressione a velocità variabile ottimizza la qualità complessiva del film anche a fronte di limiti di capacità del supporto; un'altra caratteristica interessante è la possibilità per lo spettatore di scegliere se riprodurre un film DVD con il rapporto tra lunghezza e larghezza delle immagini tipico della televisione (4 a 3) o con un rapporto simile a quello originale dei film (16 a 9).

I miglioramenti delle tecnologie digitali contribuiranno notevolmente anche alla qualità sonora dei film DVD. Come sa bene chiunque abbia visto *Jurassic Park* o *Apollo 13* in una sala cinematografica dotata di apparecchiature perfezionate, un audio multicanale ad alta fedeltà contribuisce notevolmente alla qualità dell'esperienza complessiva. Il formato DVD per i film prevede due metodi di codifica audio: la compressione AC-3 multicanale (5.1) dei Dolby Laboratories o il Pulse Code Modulation (PCM); vi è anche l'opzione dell'audio MPEG. Il sistema Dolby raggiunge una velocità di ben 448 000 bit al secondo per creare cinque canali indipendenti con qualità sonora a livello di un CD, oltre a un sesto canale dedicato a effetti sonori a bassa frequenza (da questa configurazione deriva la denominazione «5.1»). Le specifiche DVD per i dischi esclusivamente audio - i successori dei CD audio odierni - non sono state ancora completamente definite, in attesa di un contributo dell'industria discografica e di enti indipendenti come l'Audio Engineering Society.

La maggior capacità dei DVD permetterà di realizzare videogiochi, prodotti per il tempo libero e opere di consultazione didattiche più ricchi a un costo di produzione circa pari a quello dei CD attuali. I produttori di opere multimediali prevedono evidentemente che il ditale dei CD si trasformi nel secchio dei DVD. Invece di sprecare tempo a limare le loro creazioni per ridurle a soli 680 megabyte, gli autori di opere multimediali avranno presto il lusso di doversi chiedere come riempire di materiale divertente i DVD a due strati da 8,5 gigabyte di capacità, o addirittura quelli a quattro strati da 17 gigabyte.

Versioni di lettura/scrittura dei *drive* DVD dovrebbero uscire sul mercato nel 1998, e sia il DVD-R (scrivibile una sola volta) sia il DVD-RAM (che si può anche cancellare) promettono di essere molto più potenti e utili dei formati odierni CD-R e CD-E (*erasable*, cancellabile). Fino a oggi i sistemi ottici di registrazione hanno fatto uso soprattutto di tecnologie magneto-ottiche, ma per i DVD-RAM è più probabile che il mezzo di registrazione sia costituito dai cosiddetti materiali a cambiamento di fase. In questa ipotesi, si deposita sulla superficie del substrato RAM una pellicola policristallina sottile a grana finissima.

Un formato per tutti

All'inizio del 1995 sembrava improbabile che si riuscisse a definire in tempi brevi un singolo approccio alla seconda generazione di compact disc. Due gruppi di produttori di elettronica di consumo - uno guidato dalla Sony e dalla Philips (co-inventori del CD) e un altro capitanato dalla Toshiba, dalla Matsushita e dalla Time Warner - avevano elaborato indipendentemente un proprio progetto. La Sony e la Philips proponevano il *multimedia compact disc* (MMCD), capace di immagazzinare 3,7 gigabyte di dati su un disco a una sola facciata molto simile a un CD attuale. Ma il gruppo della Toshiba aveva in programma un formato radicalmente nuovo, detto SD (da «super densità»), capace di contenere cinque gigabyte su ciascun lato di un disco a due facciate.

All'inizio la Toshiba si guadagnò l'appoggio di molti operatori del settore cinematografico, convinti che il formato SD fosse l'unico a poter contenere un lungometraggio digitale con colonna sonora di qualità paragonabile a un CD audio. Altri però si sentivano più a proprio agio con l'MMCD, un'estensione prudente del ben collaudato, anche se ormai sfruttato al limite, formato CD. Ma né gli studi cinematografici, né le società produttrici di calcolatori volevano due formati incompatibili, con il rischio che si ripetesse la catastrofe dello scontro tra VHS e Betamax nella videoregistrazione. Una contesa avrebbe prima rallentato gravemente la diffusione di entrambi i formati finché uno dei due non avesse avuto il sopravvento, e poi danneggiato i consumatori sfortunati che avessero investito nel formato perdente quando i fornitori di contenuti informativi fossero passati a quello vincitore.

Un gruppo di esperti di registrazione di dati, costituito inizialmente dalla Sony e dalla Philips per valutare il loro formato, ampliò l'ambito del proprio intervento in modo da considerare anche la proposta della Toshiba. In qualità di *chairman* di questo gruppo, mi sono trovato al centro delle discussioni, che si sono protratte per tutto il 1995 e la prima metà del 1996. Fin dall'inizio il Technical Working Group (TWG) aveva chiarito quale fosse il suo obiettivo e la sua raccomandazione fondamentale: giungere a un unico formato capace di adattarsi a tutta una famiglia di tipi di dischi altamente compatibili, sia per sola lettura, sia registrabili.

Nell'agosto 1995 il TWG considerò entrambi i formati accettabili per le applicazioni informatiche, ma ritenne inammissibile che non fossero stati unificati. Divenne evidente che alcune società di informatica erano già pronte a scegliere uno dei due formati, e ben presto si iniziò a parlare seriamente di combinare gli elementi migliori di entrambi i progetti. Gli sponsor dei due formati alla fine accettarono la proposta dell'IBM, che si metteva a disposizione per mediare le divergenze tecniche, e il 15 settembre giunsero a un accordo su un solo formato, che mutuava il substrato sottile e il codice di correzione degli errori dalla proposta Toshiba/Time Warner, e il codice di modulazione del segnale dal formato Sony/Philips.

L'8 dicembre vennero definiti i rimanenti particolari del formato unificato, con l'annuncio della nascita del nuovo formato per i dischi a sola lettura, battezzato DVD o *digital versatile disc*. Il TWG ha poi continuato a lavorare con le società della DVD Alliance, concentrandosi sugli aspetti tecnici del DVD-RAM (cancellabile) e del DVD-R registrabile (scrivibile una sola volta). Oggi l'obiettivo del TWG è di incoraggiare le società che hanno adottato il formato DVD a massimizzare il livello di compatibilità tra i mezzi DVD-ROM, DVD-RAM e DVD-R e i loro prodotti, anche in ambienti applicativi diversi quali i personal computer e i sistemi per il tempo libero.

Un confronto tra DVD e CD

CARATTERISTICA	NUOVO FORMATO	VECCHIO FORMATO
Diametro del disco	120 millimetri	120 millimetri
Struttura del disco	Due substrati di 0,6 millimetri di spessore ciascuno	Un substrato di 1,2 millimetri di spessore
Dimensione minima delle buche	0,4 micrometri	0,83 micrometri
Lunghezza d'onda del laser	Da 635 a 650 nanometri	780 nanometri
Capacità	Due strati, uno per facciata, per un totale di 9,4 gigabyte Due strati, sulla stessa facciata, per un totale di 8,5 gigabyte Quattro strati, due per facciata, per un totale di 17 gigabyte	Uno strato, su una sola facciata, per un totale di 0,68 gigabyte
Apertura numerica	0,60	0,45
Densità delle tracce	13 400 tracce per centimetro	6300 tracce per centimetro
Densità dei bit (lineare)	37 800 bit per centimetro	17 000 bit per centimetro
Velocità di trasferimento dei dati	11 megabit al secondo	Da 1,2 a 4,8 megabit al secondo
Densità dei dati	0,5 gigabit per centimetro quadrato	0,1 gigabit per centimetro quadrato

Lisa Burnett

Per registrare ciascun bit si usa un impulso laser intenso ma istantaneo, che fonde una regione della pellicola di dimensioni inferiori al micrometro. Dato che un'area così piccola si raffredda molto rapidamente, la zona fusa non riesce a ricristallizzare, ma rimane «congelata» in uno stato amorfo, disordinato, che, fra le altre proprietà, riflette la luce assai meno della fase cristallina. Questa differenza di riflettività permette a un fascio di lettura a bassa intensità - che non causa la rifusione del materiale - di decodificare i dati.

Una serie di intense ricerche ha consentito di identificare e mettere a punto materiali abbastanza sensibili per registrare i dati, abbastanza stabili per conservarli e abbastanza durevoli per resistere a centinaia di migliaia di cicli di registrazione e cancellazione. Oltre a ciò, la registrazione tramite cambiamento di fase presenta numerosi altri vantaggi. Dato che la riproduzione dipende solo dall'intensità della luce riflessa dal disco, si possono usare le stesse testine di lettura ottiche che si impiegano nei lettori di DVD-ROM. La riproduzione delle registrazioni magneto-ottiche richiede invece componenti molto diversi, capaci di rilevare piccole variazioni della polarizzazione del fascio di lettura riflesso. Può darsi quindi che, diversamente da quanto accade oggi per i CD, il costo di una unità DVD-RAM che si basa sulla tecnologia a cambiamento di fase non risulti molto superiore a quello dei suoi analoghi a sola lettura. Inoltre i dischi DVD-RAM si potranno riprodurre anche nei lettori DVD-ROM standard, ammesso che tutti gli altri aspetti del formato DVD-RAM vengano progettati in modo da rispettare questa compatibilità.

Oggi esiste una varietà di dispositivi asportabili per la registrazione di dati da confondere i consumatori: dischetti ad alta e bassa capacità, dischi rigidi asportabili in cartuccia, mezzi magneto-ottici e vari tipi di cartucce a nastro magnetico. Nessuno prevede che i DVD-R e i DVD-RAM registrabili del futuro sostituiscano i dischi rigidi magnetici, che sono superiori per prestazioni e diffusissimi, ma sicuramente questi formati hanno le potenzialità per offrire una soluzione unica a tutte le applicazioni che richiedono mezzi asportabili e accessibili sia in lettura sia in scrittura. Considerando che il costo dei DVD scrivibili dovrebbe scendere a 30-75 lire per megabyte - un valore abbastanza basso da soddisfare anche il più impenitente «scaricatore di dati» da Internet - le unità per DVD registrabili promettono capacità elevata, buone prestazioni nell'accesso diretto e basso costo, oltre alla compatibilità con il formato che dovrebbe diventare lo standard più diffuso per la distribuzione di informazioni, ossia il DVD-ROM. Un solo apparecchio DVD-RAM potrebbe quindi gestire molte applicazioni del personal computer, dall'editoria personale alla realizzazione di opere multimediali, dalla distribuzione di contenuti informativi al trasferimento, all'archiviazione e alla copia di sicurezza dei dati.

Il formato DVD è stato ideato tenendo conto delle estensioni che saranno sviluppate in futuro. Per esempio, la messa a punto di laser affidabili di piccola lunghezza d'onda, che emettano luce verde o blu, potrebbe forse raddoppiare ancora la densità dei dati (si veda l'articolo *Una tecnologia per CD a laser blu* di Robert L. Gunshor e Arto V. Nurmikko a pagina 70). Si può immaginare che un giorno derivati della tecnologia DVD di base riescano ad ammassare 50 gigabyte o più su un solo disco da 1,2 millimetri di spessore, in sostanza una piccola biblioteca su un singolo supporto.

Un piccolo esercito di esperti sta attualmente completando le specifiche generali della famiglia unificata di formati DVD. Essi devono cercare di fare scelte tecniche che offrano la base più funzionale per le applicazioni future, compresi numerosi impieghi ancora da immaginare. Proprio come la televisione si è evoluta rapidamente in un veicolo molto diverso da una «radio con immagini», così anche le applicazioni basate sul nuovo formato di compact disc si svilupperanno probabilmente in direzioni sorprendenti e imprevedibili per molti anni a venire.

ALAN E. BELL, laureatosi all'Imperial College dell'Università di Londra, inizia la sua carriera nel 1974 alla RCA, progettando diversi dispositivi utilizzati negli attuali prodotti per la registrazione digitale. Nel 1982 passa all'IBM e - dal 1995 - è coordinatore del Technical Working Group, un gruppo di rappresentanti delle più importanti società di informatica che ha contribuito a unificare i diversi formati nel DVD.

POHLMANN KEN C., *The Compact Disc Handbook*, The Computer Music and Digital Audio Series, 5, A-R Editions, Madison, 1992.

Una FAQ riguardante i dischi digitali versatili è consultabile presso l'indirizzo Internet http://www.ima.org/forums/imf/dvd/faq.html

Una tecnologia per CD a laser blu

Indurre emissione di luce blu da cristalli semiconduttori non è facile, ma i benefici per le applicazioni multimediali sarebbero enormi

di Robert L. Gunshor e Arto V. Nurmikko

Fino a poco tempo fa i laser blu a semiconduttore erano estremamente difficili da realizzare. Per valutarne in laboratorio le potenzialità, alcuni studiosi si sono affidati a una soluzione transitoria in cui un cristallo ottico non lineare converte la lunghezza d'onda di un laser infrarosso a semiconduttore. In futuro fasci laser blu dovrebbero essere prodotti direttamente da minuscoli cristalli di semiconduttore con ampia banda proibita.

IBM Almaden Research Center

Il dispositivo fondamentale all'interno di un lettore di compact disc o di CD-ROM è un laser piccolo ma potente che, funzionando come una puntina di giradischi eccezionalmente fine, permette all'apparecchio di leggere le informazioni registrate sulla superficie del CD in forma di minuscole buche. (Nella prossima generazione di riproduttori di dischi ottici il laser, oltre a leggere, scriverà sul disco impartendo al materiale modifiche reversibili.) È la lunghezza d'onda della luce laser a limitare il numero di buche - e quindi la quantità di dati - che si può registrare sul compact disc: quanto più la lunghezza d'onda è piccola, tanto più sono piccole le buche che l'apparecchio è capace di leggere.

Oggi i laser dei lettori di CD sono fatti con arseniuro di gallio (GaAs) e semiconduttori a esso affini; si tratta di composti che, dopo essere stati portati in uno stato di alta energia, emettono luce con una lunghezza d'onda di circa 820 nanometri. Questa luce infrarossa riesce a leggere buche di dimensioni non inferiori al micrometro, equivalenti a circa un cinquantesimo del diametro di un capello umano. Come spiega Alan E. Bell nell'articolo a pagina 64, si stanno elaborando nuovi formati di dischi ottici capaci di sfruttare i grandi progressi dei laser rossi a semiconduttore per aumentare la densità delle informazioni. Ma i laser a diodo blu - che emettono luce a una lunghezza d'onda di 460 nanometri - potrebbero andare ben oltre, essendo in grado di leggere buche molto più piccole. Un CD audio dotato di buche così piccole potrebbe, per esempio, contenere tutte e nove le sinfonie di Beethoven invece di una sola. Anche le applicazioni multimediali ne trarrebbero benefici enormi.

Nonostante le loro potenzialità, però, i laser blu per lettori di CD si sono rivelati difficili da realizzare. Per capirne la ragione occorre considerare come funzionano questi dispositivi a semiconduttore. I laser all'interno di un lettore di CD sono minuscoli cristalli pressoché privi di imperfezioni, ottenuti con tecniche di deposizione atomica, che fungono da diodi e sono divisi in due regioni fondamentali. Ogni lato di questi diodi ha caratteristiche elettriche molto specifiche. Sul lato di tipo *n* esiste una sovrabbondanza di elettroni conduttori di corrente, mentre sull'altro lato, di tipo *p*, questa funzione è svolta da «buche» - cioè particelle di carica positiva che rappresentano l'assenza di un elettrone - presenti in numero ugualmente grande. Se si applica una tensione positiva al lato *p* e una negativa al lato *n*, buche ed elettroni si muovono le une verso gli altri.

Queste particelle si incontrano in una trincea estremamente sottile, una terra di nessuno detta pozzo quantico, dove si ricombinano annichilandosi a vicenda e, in condizioni opportune, emettendo fotoni. Quando a questa emissione si affianca un meccanismo di retroazione - costituito da una coppia di specchi altamente riflettenti incorporata nel dispositivo che «ricicla» i fotoni - si ottiene un'emissione laser: tutti i fotoni prodotti risultano coerenti e conferiscono al sottilissimo fascio emesso le sue qualità eccezionali.

L'energia dei fotoni, che determina la lunghezza d'onda della luce prodotta, dipende a sua volta da uno dei parametri fondamentali dell'elettronica: l'energia della banda proibita, ovvero, approssimativamente, l'«elasticità elettronica» del semiconduttore in cui si verifica il fenomeno di ricombinazione (*si veda l'illustrazione a pagina 73*). Per i laser a diodo di arseniuro di gallio l'energia della banda proibita è di circa 1,45 elettronvolt; per produrre luce blu questo valore deve essere quasi raddoppiato. Occorre quindi ricorrere a un'altra categoria di semiconduttori, dotati di banda proibita ampia: fra essi vi sono il seleniuro di zinco (ZnSe) - un composto II-VI, così chiamato per la collocazione nel sistema periodico degli elementi che lo compongono - e il nitruro di gallio (GaN), un composto III-V. Questi materiali potrebbero essere appropriatamente definiti semisolanti, nome che evidenzia immediatamente uno dei problemi fondamentali che si incontrano nell'impiegarli in un dispositivo elettrico.

All'inizio degli anni ottanta un gruppo di scienziati tentò di formare cristalli di semiconduttori II-VI con l'epitassia a fascio molecolare. Il metodo consiste nel far deporre delicatamente su un substrato gli elementi chimici che interessano da una sorta di «doccia» atomica all'interno di una camera a vuoto ultraspinto, creando un «grattacielo» atomico. (Il processo è simile alla costruzione di un edificio mattone su mattone, in questo caso con una precisione a livello atomico.) Con questa tecnica gli scienziati realizzarono presto materiali speciali di qualità cristallina elevata, e i loro studi permisero di capire meglio come funzioni un pozzo quantico e come ottenere luce blu e verde da materiali artificiali estremamente raffinati. (La lunghezza d'onda della luce verde è solo di poco superiore a quella della luce blu.) Ma la ricerca di un laser a diodo blu o verde utilizzabile in pratica non era finita. Questi primi dispositivi emettevano luce solo se ricevevano energia da un altro laser «di eccitazione» che doveva avere dimensioni molto maggiori.

Per quasi dieci anni non si riuscì a trovare un altro modo per alimentare di energia i laser verdi e blu. Era facile drogare cristalli di ZnSe in modo che avessero un eccesso di elettroni, ma creare una sovrabbondanza di buche era molto più problematico; non si riusciva quindi a produrre una giunzione *p-n*. Poi, nel 1990, si perfezionò un sistema per utilizzare atomi di azoto nel processo di drogaggio, riuscendo finalmente a introdurre un eccesso di buche nel seleniuro di zinco. L'estate successiva, ricercatori della 3M e, indipendentemente, il nostro gruppo della Brown University e della Purdue University descrissero i primi laser a diodo blu e verdi. Questi laser non potevano funzionare in continuo e andavano raffreddati alla temperatura dell'azoto liquido (77 kelvin). Poco dopo, però, entrambi i gruppi riuscirono a migliorare i propri dispositivi ottenendo sia un funzionamento in continuo a 77 kelvin sia un funzionamento a impulsi a temperatura ambiente.

Nell'autunno del 1993 studiosi dei laboratori Sony e il gruppo Brown-Purdue realizzarono un laser a diodo capace di funzionamento continuo fino a qualche decina di secondi, a temperatura ambiente e alimentato da una tensione di soli cinque volt. I primissimi laser verdi e blu potevano richiedere tensioni di ben 30 volt. Il drastico miglioramento era dovuto ad abili operazioni di ingegneria atomica dei materiali. Molto recentemente la Sony ha creato un laser a diodo verde che emette luce a 520 nanometri e funziona continuamente per circa 100 ore a temperatura ambiente. Il gruppo Brown-Purdue, dal canto suo, ha presentato un laser blu con lunghezza d'onda di 460 nanometri.

Naturalmente 100 ore non bastano ancora per un impiego pratico, ma, sapendo che i laser infrarossi a base di GaAs oggi diffusissimi hanno all'inizio sofferto di problemi sostanzialmente identici, ci manteniamo ottimisti. Ultimamente ricercatori delle Nichia Chemical Industries, in Giappone, hanno fatto progressi lavorando con materiali a base di nitruro di gallio. Questi studiosi hanno realizzato eccellenti diodi che emettono luce blu e hanno compiuto dimostrazioni preliminari di un laser blu, sia pure in condizioni di tensione elettrica elevata e funzionamento a impulsi.

Nonostante le loro limitazioni, i laser a diodo blu e verdi si possono migliorare in vari modi. Regolando la disposizione degli atomi nel cristallo si possono realizzare cristalli più resistenti alle alte densità di corrente necessarie per alimentare di energia il laser. L'illustrazione a pagina 72 mostra lo schema

di un laser realizzato con una combinazione di semiconduttori II-VI, ricavati per lo più dal seleniuro di zinco. La successione degli strati depositati con estrema precisione ottimizza il flusso della corrente elettrica verso il pozzo quantico, uno strato ultrasottile di seleniuro di cadmio e zinco (ZnCdSe), che è il cuore ottico ed elettronico del laser. Questa configurazione offre inoltre una sede ottica adeguata per guidare i fotoni emessi e produrre il meccanismo di retroazione. Il pozzo quantico, di spessore pari a circa un centesimo di micrometro, si forma confinando elettroni e buche tra i due strati di seleniuro di zinco e zolfo (ZnSSe), che hanno un'energia della banda proibita alquanto grande.

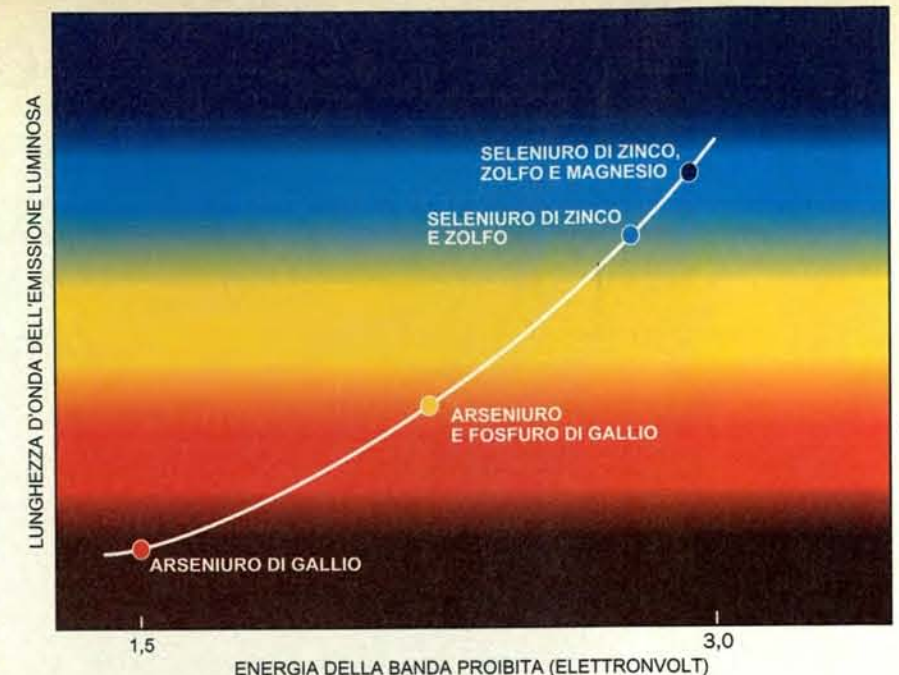
La ricombinazione di un elettrone e di una buca in un semiconduttore può sembrare abbastanza intuitiva, ma la meccanica quantistica priva queste particelle della propria individualità. In effetti gli elettroni e le buche subiscono attrazione elettrostatica e formano coppie dette eccitoni. Di norma, a temperatura ambiente, le vibrazioni all'interno del reticolo cristallino scindono le coppie; la stessa cosa accade quando si confinano molte particelle in un'area ristretta: gli urti tra le particelle tendono a separare gli eccitoni. Ma in uno strato sottile di semiconduttore con ampia banda proibita, come il ZnCdSe, le coppie eccitoniche sono così strette da rimanere unite anche a temperatura ambiente o in un affollatissimo dispositivo laser. Per queste coppie a vita lunga l'emissione di fotoni risulta più probabile, sicché il dispositivo richiede meno corrente per mantenere nel tempo l'emissione laser.

Un'altra caratteristica di progetto serve a correggere un problema che insorge perché la lunghezza d'onda delle emissioni blu e verde è maggiore di quasi due ordini di grandezza dello spessore del pozzo quantico. A questo scopo si utilizza un altro materiale, il seleniuro di zinco, magnesio e zolfo (ZnMgSSe), per definire una guida d'onda ottica che serve a intrappolare la luce in senso verticale.

Nel 1993 alcuni ricercatori dei Philips Laboratories dimostrarono come questo espediente migliori il funzionamento del laser. Per ottenere la guida d'onda si rimuove fisicamente materiale ai lati del cristallo mediante tecniche litografiche, creando così un profilo ad «altopiano». L'intera struttura incanala quindi la luce laser soprattutto in una direzione, quella assiale; la radiazione rimbalza tra le due facce terminali tagliate lungo un piano cristallino - che costituiscono specchi quasi perfetti - poi esce da una delle due. Una configurazione analoga basata su nitruri III-VI ha permesso recentemente di ottenere emissione laser blu da questi semiconduttori; in questo caso il pozzo quantico è composto da nitruro di indio e gallio (InGaN), mentre lo strato di confinamento elettronico e quello di confinamento ottico sono costituiti rispettivamente da GaN e da nitruro di gallio e alluminio (AlGaN).

Per produrre luce blu (*in basso*) i cristalli di semiconduttore devono essere collocati in maniera molto precisa. Due strati sottili di seleniuro di zinco e zolfo (*in verde*) ne affiancano uno di seleniuro di zinco e cadmio (*in blu*). Lo strato superiore di ZnSSe reca un eccesso di buche, mentre quello inferiore contiene un eccesso di elettroni. Se si applica una tensione agli strati esterni del cristallo queste particelle fluiscono nello strato di ZnCdSe, dove si annichilano reciprocamente emettendo fotoni (*qui a destra*). Strutture di seleniuro di zinco, magnesio e zolfo (*in giallo*) formano una guida d'onda ottica. Lo strato inferiore di ZnMgSSe contiene un eccesso di elettroni e confina con un substrato di arseniuro di gallio collegato a una sorgente di tensione negativa. Sopra lo strato di ZnMgSSe contenente un eccesso di buche giacciono strati alternati di seleniuro e di tellururo di zinco, che sono collegati a una sorgente di tensione positiva.

La parte superiore della struttura, al di sopra dello strato tipo *p* di ZnSSe, è chiamata contatto di Zn(Se,Te) a separazione decrescente ed è ottenuto alternando strati di seleniuro di zinco e di tellururo di zinco. Uno dei problemi dei primi laser a diodo blu e verdi era che, pur potendo funzionare in continuo a temperatura ambiente, si surriscaldavano in meno di un secondo se li si collegava direttamente a un alimentatore. Il riscaldamento avveniva a livello del contatto elettrico, dove l'elettrodo di metallo toccava il semiconduttore a drogaggio *p*. La cospicua incompatibilità elettronica, o barriera di Schottky, tra i due materiali ostacolava la conduzione. Si dovette attendere fino alla fine del 1992 perché i ricercatori del gruppo Brown-Purdue e, indipendentemente, quelli della Sony trovassero un modo di aggirare il problema. La soluzione si basava su una lega a gradazione progressiva di ZnSe e ZnTe, che sostituisce alla netta barriera di conduzione tra metallo e ZnSe un passaggio più dolce.

Oggi gran parte della ricerca in tutto il mondo si concentra sul tentativo di prolungare la vita operativa di un laser a diodo blu o verde a semiconduttore II-VI funzionante in continuo, e sul miglioramento della qualità dei nitruri utilizzati. Oggi il principale problema riguardante i materiali a base di ZnSe - che per tutti gli altri aspetti rappresentano un pezzo di bravura straordinario dell'attuale fisica dello stato solido - è un difetto cristallino che insorge quando si fanno nucleare i primi strati di ZnSe su un substrato di GaAs, che ha un reticolo cristallino dai parametri molto simili. A volte alcuni atomi (anche solo uno su circa un milione) si allineano erroneamente e inducono l'accrescimento di un piano cristallino in soprannumero. Questo difetto di impilamento si espande via via che il cristallo cresce; quando attraversa il pozzo quantico, esso fornisce agli elettroni e alle buche siti dove possono ricombinarsi in maniera non radiativa, cioè senza produrre un fotone. Ciascuno di questi eventi non radiativi libera nel reticolo circostante circa 2,5 elettronvolt di energia, una quantità paragonabile a quella che tiene unito il cristallo stesso. Questo fenomeno causa spesso la formazione di ulteriori difetti, che accrescono l'accumulo di calore fino a portare al cedimento del laser.

Svariate tecniche diagnostiche, come la microscopia ottica ad alta risoluzione e la microscopia elettronica a trasmissione, hanno rivelato almeno in parte come avvenga questo processo in un laser a diodo semiconduttore II-VI. Basandosi su queste informazioni, i principali laboratori stanno valutando una serie di metodi per prevenire i difetti di impilamento, che fanno ricorso alle numerose tecniche disponibili per trattare il materiale laser in modo da accrescerne la resistenza. Con i nitruri le difficoltà sono di altro tipo: questi materiali sono molto tenaci a causa della configurazione dei legami chimici al loro interno, e quindi è meno probabile che un'intensa corrente generi difetti di impilamento. Attualmente, però, i difetti nei nitruri tendono a essere tanto numerosi da impedire che si attivi l'emissione laser. Eppure le soluzioni di questi problemi non devono essere lontane. Considerate le loro potenzialità per applicazioni di registrazione ottica ad alta densità e per una serie di altri usi, dalle trasmissioni televisive alla diagnostica medica, i laser a base di semiconduttori con ampia banda proibita dovrebbero avere un «luminoso» futuro.

È l'energia della banda proibita di un cristallo di semiconduttore a determinare l'energia dei fotoni che esso emette e quindi la lunghezza d'onda della luce così prodotta. Un materiale con una banda proibita ampia, come il seleniuro di zinco e zolfo, emette luce di piccola lunghezza d'onda, nella regione blu-verde dello spettro.

ROBERT L. GUNSHOR e ARTO V. NURMIKKO collaborano da ormai 12 anni. Gunshor è professore di microelettronica alla Purdue University, mentre Nurmikko insegna ingegneria e fisica alla Brown University, dove dirige il Centro di ricerca sui materiali avanzati.

NURMIKKO A. V. e GUNSHOR R. L., *Physics and Device Science in II-VI Semiconductor Visible Light Emitters* in «Solid State Physics» (Academic Press), vol. 49, pp. 205-282, 1995.

AGRAWAL GOVIND, *Semiconductor Lasers: Past, Present and Future*, American Institute of Physics Press, 1995.

Modelli a quark e QCD

Sebbene la teoria esatta, la QCD, sia assai complicata, un modello semplice della struttura a quark dei protoni spiega bene molte loro proprietà. Dopo trent'anni se ne capisce la ragione, anch'essa assai semplice

di Giacomo Morpurgo

La curiosità di sapere come è fatta la materia è antica quanto l'uomo. Appartiene alla classe di quelle curiosità che ci accompagnano da sempre: come è fatto l'universo, come è nata la vita, come funziona il cervello.

La ricerca dei costituenti della materia è una storia lunga; qui ricorderò soltanto che, ancora all'inizio di questo secolo, molti eminenti scienziati dubitavano dell'esistenza reale degli atomi. Questo secolo ha segnato la nascita della «fisica delle particelle». E, per entrare subito in argomento, vediamo di riassumere a grandi linee l'evoluzione della situazione per arrivare poi al quadro attuale. In primo luogo occorre ricordare qualche nome.

Gli adroni (dal greco αδρος, cioè «robusto») sono quelle particelle, come il protone, il neutrone o il pione, che interagiscono fortemente tra loro. Per esempio, protoni e neutroni si legano fortemente a formare i nuclei atomici. Invece i leptoni (da λεπτος, «debole, leggero») interagiscono solo debolmente o elettricamente. Essi comprendono l'elettrone, il muone, e il neutrino. All'inizio di questa storia (sul finire degli anni quaranta) gli adroni e i leptoni fin qui elencati (più il fotone, cioè il quanto di luce) erano le sole particelle conosciute. Essendo poche, si pensava che fossero tutte elementari; un'etichetta, quella di «elementare», in continua evoluzione.

È ovvio che se esistessero solo uno o due atomi tenderemmo a pensarli elementari (anche se poi non lo sono). Se però ne esistono cento tipi, è naturale immaginare che non tutti siano elementari. Da tempo sappiamo che gli atomi sono costituiti da un nucleo e da elettroni. A loro volta, i nuclei sono fatti di protoni e neutroni e la loro varietà è dovuta al diverso numero di protoni e neutroni che contengono. Che dire degli adroni?

Poiché alla fine degli anni quaranta gli unici adroni noti erano i «nucleoni» (cioè protone e neutrone) e i pioni, i più li ritenevano elementari. Si poteva leggere in testi ormai classici: la forza tra due protoni deve dipendere semplicemente dalla loro distanza, perché i protoni sono elementari, dunque semplici. Pensando che pioni e nucleoni fossero elementari, si credeva che, una volta risolte (arduo problema) le equazioni quantistiche del sistema pione-nucleone, la fisica delle interazioni forti sarebbe stata un capitolo chiuso. Quest'idea prevaleva nel decennio 1950-1960. Rimane, anche dopo Galileo, una tendenza a credersi al centro del mondo (temporalmente); molti fisici credevano che proprio a loro fosse toccata la sorte di aver trovato, nel giro di pochi anni, tutti gli adroni; talché, risolti certi problemi matematici, tutto sarebbe stato chiaro.

Gli adroni strani

Nel decennio 1950-1960, mentre la posizione prevalente era quella descritta, si andavano scoprendo molti nuovi adroni; furono detti «strani», per via di certe curiose proprietà. Si capì poi che queste proprietà si potevano spiegare attribuendo a ogni adrone un numero intero S (detto «stranezza»). I vecchi adroni (pioni e nucleoni) avevano $S = 0$, mentre gli adroni strani hanno S diverso da zero.

Anche se S non ha nulla a che fare con la carica elettrica Q, vi è qualche analogia. Come per la carica, la stranezza totale di due adroni è la somma delle due stranezze. La stranezza si conserva in una reazione forte; cioè se A e B, urtandosi, si trasformano in C e D, la somma delle stranezze di A e B eguaglia la somma di quelle di C e D. Se A e B sono pioni e/o nucleoni (cioè adroni con $S = 0$) e se l'adrone C prodotto in $A + B \rightarrow C + D$ è strano, anche D deve esserlo, con stranezza eguale e di segno opposto a quella di C. Il fatto che in un urto di pioni e nucleoni un adrone strano fosse sempre prodotto assieme a un altro strano era stato proprio uno dei motivi del termine «strano» attribuito a questi adroni. A differenza della carica elettrica, però, che è sempre conservata, la stranezza non lo è in certi processi deboli; perciò gli adroni strani, che hanno una vita media di circa un miliardesimo di secondo, decadono spontaneamente in particelle più leggere.

Enrico Fermi e gli stati eccitati del protone

Sempre negli anni cinquanta, oltre a scoprire gli adroni strani si trovò che ogni adrone, non strano o strano, aveva diversi livelli energetici, proprio come gli atomi o i nuclei. Chi capì questo fatto molto presto fu Enrico Fermi. Tra il 1952 e il 1953, a Chicago, egli aveva studiato gli urti di pioni con protoni, una delle prime esperienze del genere con un acceleratore di particelle. Fermi voleva sapere come interagivano pioni e protoni, e si aspettava qualcosa di semplice, data la presunta «elementarità» di pioni e protoni. Si attendeva che il numero di pioni diffusi dai protoni variasse blandamente al variare dell'energia a cui avveniva l'urto. Invece osservò un rapido aumento della diffusione dei pioni quando la loro energia si avvicinava a un certo valore E^*, e una diminuzione oltrepassando E^*. Questo e altri fatti indicavano che all'energia E^* si formava uno stato risonante di pione e protone, cioè uno stato più stabile di quelli con energia diversa.

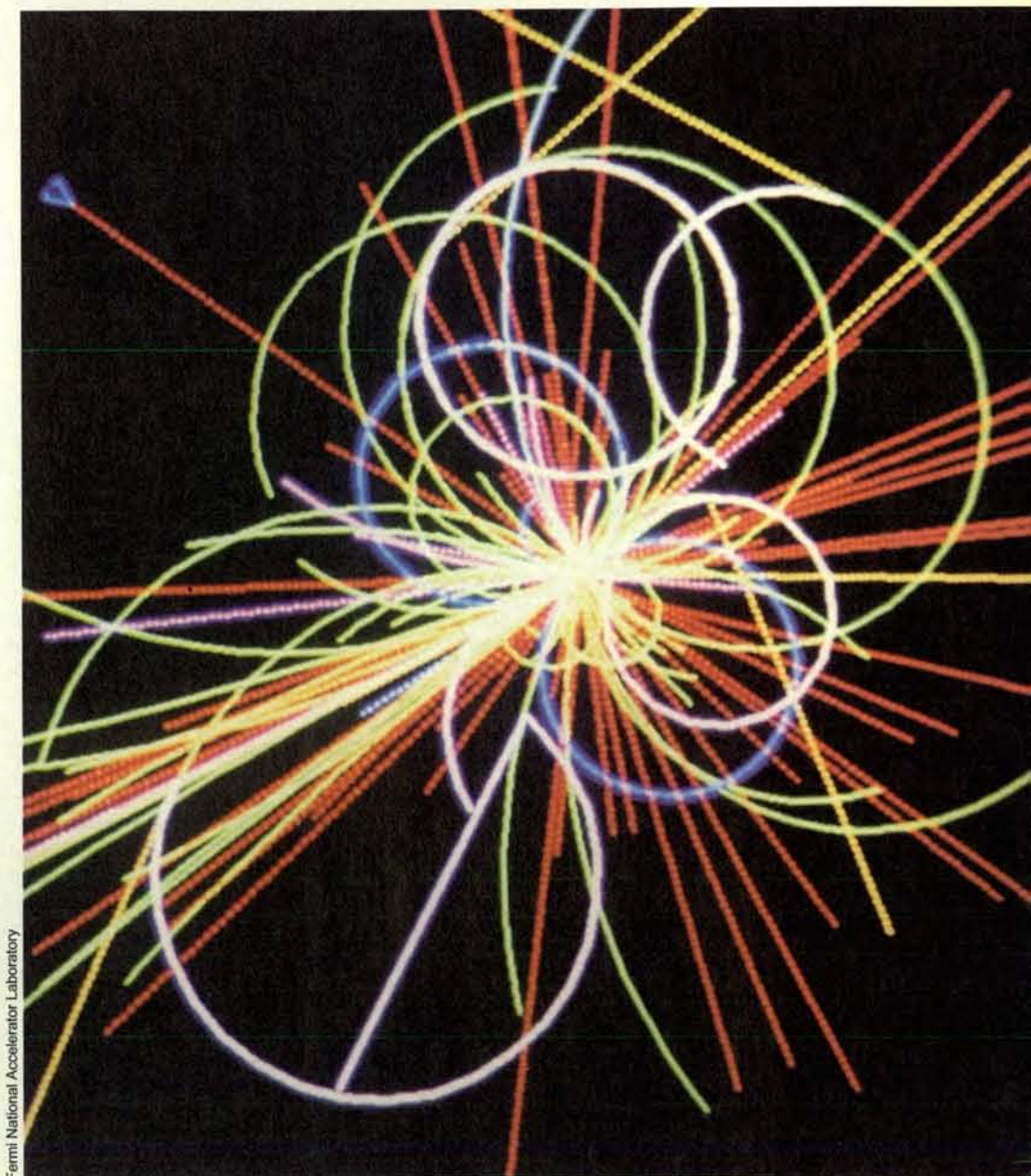
Un tale fenomeno, piuttosto comune negli urti tra sistemi complessi (per esempio tra nuclei), è inaspettato in urti tra sistemi semplici. Fermi non poté che prender atto del risultato, ma ne fu sorpreso. Ricordo, parola per parola, una frase di Fermi durante una passeggiata con lui ed Enrico Persico a Vigo di Fassa nell'agosto del 1954, tre mesi prima della sua morte: «Ho studiato l'urto pione-protone e mi accorgo che è come se avessi studiato l'urto tra due nuclei complicati». Intendeva dire che pione e protone erano complicati: il protone aveva uno stato eccitato, come quelli dei nuclei. Tale stato, una volta eccitato, durava pochissimo (circa 10^{-23} secondi), ma non per questo era meno reale. Ascoltando Fermi a Vigo, ricordai quanto mi aveva detto due anni prima a Chicago: «I risultati importanti, quelli che hanno qualche probabilità di restare, si devono poter spiegare in dieci minuti». Anche se Fermi non aveva precisato il tipo di interlocutore, il suo era un risultato «spiegabile in dieci minuti».

I quark

L'aumento del numero degli adroni conosciuti è riassunto nella figura in alto a pagina 76, che dà la situazione nei primi anni sessanta. Ciascun livello rappresenta un diverso adrone. Per ognuno di essi sono indicate alcune quantità: in primo luogo il suo «nome», un simbolo come Σ, Λ, π o K. Con Σ (1192) si intende che la massa dell'adrone Σ è 1192 in certe unità (quelle stesse in cui la massa dell'elettrone è circa 0,5 e quella del protone 938). Se sopra il livello, come nel caso del Σ, si trovano specifiche come +, 0, – vuol dire che il Σ esiste con carica +1, zero, oppure –1. (Questi tre Σ hanno masse che differiscono di qualche unità; 1192 è la massa media.) Sotto il livello è riportato il valore della stranezza S. Infine ho separato gli adroni in due gruppi: a sinistra i mesoni, che hanno spin J intero (in termini della classica trottola, J rappresenta il momento angolare della particella); a destra i barioni, con J semintero. I mesoni (o i barioni) su una stessa colonna hanno tutti lo stesso spin, indicato in cima alla colonna.

La prima cosa da notare è che, nei primi anni sessanta, si conoscevano già molti adroni: nove mesoni con $J = 0$ e nove con $J = 1$; quanto ai barioni, otto con spin 1/2 e dieci con spin 3/2. A proposito di questi, l'Ω^- fu previsto da Murray Gell-Mann prima di essere scoperto nel 1964; la sua scoperta (*si veda l'illustrazione a pagina 78*) fu un evento straordinario, e costituì il primo passo per l'ingresso dei quark nella fisica.

Il più basso livello dei mesoni è il pione, π (140), nei suoi tre stati di carica. Per i barioni con $J = 1/2$ il livello più basso è il nucleone N (939) nei suoi due stati di carica protone, p, e neutrone, n. Tra i barioni con $J = 3/2$ lo stato più basso è l'oggetto individuato da Fermi, l'N^*(1232), in quattro stati di carica. Oggi gli adroni conosciuti sono molti di più di quelli dell'illustrazione; ci vuole un quaderno per registrarli tutti.

Dunque dato il loro numero, occorreva pensare che gli adroni non fossero elementari e ci si chiese quali ne fossero i costituenti. Gell-Mann ipotizzò questi costituenti nel 1964, li chiamò «quark» e li suppose dotati di certe proprietà che vedremo tra poco. Per spiegare la varietà degli adroni conosciuti (quelli, appunto, riportati nella figura in alto a pagina 76) occorreva ipotizzare tre quark, che Gell-Mann indicò con i simboli u, d, s, ciascuno con spin 1/2. L'introduzione dei tre diversi quark

Una simulazione al computer di uno scontro frontale tra due protoni di altissima energia.

MESONI

$J=0$: η' (958) 0; η (550) 0; K (494) +,0 (S = +1), 0,– (S = –1); π (140) +,0,–

$J=1$: Φ(1020) 0; K*(892) +,0 (S = +1), 0,– (S = –1); ω (783) 0; ρ (770) +,0,–

BARIONI

$J=\frac{1}{2}$ (8): Ξ (1320) –,0 (S = –2); Σ (1192) +,0,– (S = –1); Λ (1115) 0 (S = –1); N (938) +,0 (p,n)

$J=\frac{3}{2}$ (10): Ω(1670) – (S= –3); Ξ*(1530) 0,– (S= –2); Σ*(1385) +,0,– (S= –1); N*(1232) ++,+,0,–

A sinistra, le prime due famiglie di mesoni: i nove mesoni con J=0 e i 9 con J=1. Vicino al simbolo di ciascun mesone è indicato il valore della massa (in unità in cui l'elettrone ha massa ≈ 0,51). La notazione +, 0, – sopra il livello di una particella significa che essa può esistere con carica elettrica +1, 0, –1. (Per esempio π^+, π^0, π^-). Per le particelle dotate di stranezza, il valore di questa è indicato sotto lo stato. A destra, le prime due famiglie di barioni: gli otto con spin J=1/2 e i 10 con J=3/2. I simboli sono gli stessi usati per i mesoni. Il livello più basso dell'ottetto, N, è il «nucleone», nei due stati protone (p) e neutrone (n). (I diversi stati di carica di ciascun barione hanno masse che differiscono di qualche unità; per esempio il neutrone ha massa 939,56 e il protone 938,27; come nel caso dei mesoni, si sono trascurate queste differenze.)

serviva a spiegare il numero di componenti di ogni famiglia di mesoni e barioni, nonché le cariche e le stranezze di ciascun mesone e barione e i loro spin. Nella tabella in basso in questa pagina si trovano le caratteristiche dei quark necessarie per questo. I mesoni (*figura in basso a sinistra nella pagina a fronte*) sono composti da un quark e un antiquark, mentre i barioni (*figura in basso a destra nella pagina a fronte*) sono costituiti da tre quark. Introducendo i quark si è realizzato dunque un grosso risparmio: tutti gli adroni conosciuti nei primi anni sessanta, ma anche i moltissimi altri scoperti successivamente, possono essere descritti secondo lo schema delle illustrazioni in basso nella pagina a fronte, cioè come stati composti da un quark e un antiquark per i mesoni e da tre quark per i barioni. Va aggiunto però che sono stati trovati anche adroni che non possono essere descritti come aggregati dei quark u, d, s. Occorre, per spiegare tutti gli adroni conosciuti oggi, assumere che esistano anche altri quark: c (iniziale di *charm*), b (da *bottom*) e t (da *top*), anch'essi indicati nella tabella qui a fianco. Dunque i diversi tipi (o «sapori» come si usa dire) di quark sono sei.

I VARI TIPI DI QUARK			
up	u	$+\frac{2}{3}$	
down	d	$-\frac{1}{3}$	
strange	s	$-\frac{1}{3}$	stranezza = –1
charmed	c	$+\frac{2}{3}$	charm = 1
bottom	b	$-\frac{1}{3}$	beauty = 1
top	t	$+\frac{2}{3}$	

I vari tipi di quark: nella prima colonna il nome, nella seconda il simbolo, nella terza la carica elettrica (in unità del modulo della carica dell'elettrone), nella quarta le caratteristiche come stranezza e così via. Aggregati contenenti i quark c, b, t sono più pesanti degli adroni a cui ci siamo riferiti nel testo. Qui ci siamo limitati ad adroni contenenti solo quark u, d, s.

Il confinamento dei quark e la QCD

Un atomo si spezza facilmente; da esso, cioè, si possono estrarre gli elettroni, separandoli tra loro e dal nucleo. Così da un nucleo si possono estrarre i protoni e i neutroni che lo compongono. Non si riesce invece a spezzare gli adroni in modo da farne uscire i quark che li costituiscono: i quark sono «confinati». Nonostante ogni sforzo, non si è mai trovato un quark isolato. E sì che la sua osservazione è facilitata dal fatto che i quark sono le sole particelle conosciute con carica frazionaria (un terzo o due terzi di quella dell'elettrone). I quark sembrano imprigionati negli adroni. In questo senso i quark sono assai diversi dalle particelle conosciute prima.

Ma il quadro non è ancora completo. Per completarlo occorre, a quanto sappiamo oggi, supporre che ciascun quark con un dato sapore - per esempio u, oppure d, oppure s (e anche c, b, t) - possieda un grado di libertà interno (detto «colore») che può assumere tre diversi valori (giallo, rosso, blu). Anzi, i tre quark in un barione devono avere colori diversi, come è indicato nella figura in basso a destra nella pagina a fronte. Come la forza tra due elettroni è dovuta al fatto che uno dei due produce un campo elettromagnetico che esercita una forza sull'altro, così la forza tra due quark colorati è dovuta a un campo «gluonico» che i quark producono: un quark colorato produce un campo gluonico che agisce sugli altri. È proprio questo campo (e lo scambio dei suoi quanti, i gluoni, oggetti simili ai fotoni, elettricamente neutri), a produrre la forza tra i quark che ne determina il confinamento. E anche il campo gluonico è confinato; di gluoni liberi non se ne trovano.

La teoria dei quark «colorati», legati dall'interazione comune con il campo gluonico, si chiama cromodinamica quantistica (QCD, dall'inglese *Quantum Chromo-Dynamics*). È analoga all'elettrodinamica quantistica, la teoria dell'interazione tra cariche elettriche e campo elettromagnetico (ma è ben più complicata). La QCD è considerata la teoria corretta delle interazioni tra quark negli adroni, ma ha due problemi: innanzitutto non si sa dimostrare che la QCD implichi il confinamento dei quark; in secondo luogo, mentre con la QCD si sanno analizzare gli urti di adroni ad alta energia, le proprietà adroniche (ad esempio le masse di mesoni e barioni) non si sanno calcolare a causa delle complicazioni matematiche. Un inciso: da tempo si usano grossi calcolatori per trattare questo tipo di problemi (si veda per esempio l'articolo *Il calcolatore nello studio dei quark* di Donald H. Weingarten in «Le Scienze» n. 332, aprile 1996). Al momento esistono, per varie ragioni che sarebbe troppo lungo illustrare qui, opinioni diverse su questo approccio, la cosiddetta QCD sul reticolo. Accanto a un folto gruppo di entusiasti, vi è un altrettanto folto gruppo di scettici.

Per concludere questo quadro, accenno a una critica che può sorgere spontanea. Ci si può chiedere: non è un po' artificiosa una descrizione in cui si introducono tante entità che però non si possono isolare, e cioè i sei quark, ciascuno in tre colori, più i gluoni? Al momento si può solo rispondere così: finché non si troverà una descrizione diversa, il che è sempre possibile, resta il fatto che siamo in grado di capire e interpretare una massa gigantesca di dati, non solo sulle proprietà dei singoli adroni, ma anche sulle loro collisioni. Vi sono punti poco chiari, come il confinamento, ma quanto ora detto rimane. Semmai il punto più oscuro è un altro. Anche se non posso parlarne qui, c'è una corrispondenza tra i quark e i leptoni (elettrone, muone, leptone τ e neutrini), importante per la validità della teoria, di cui non si conosce l'origine profonda. Finché ciò non sarà chiarito, qualcosa di insoddisfacente resterà sul fondo.

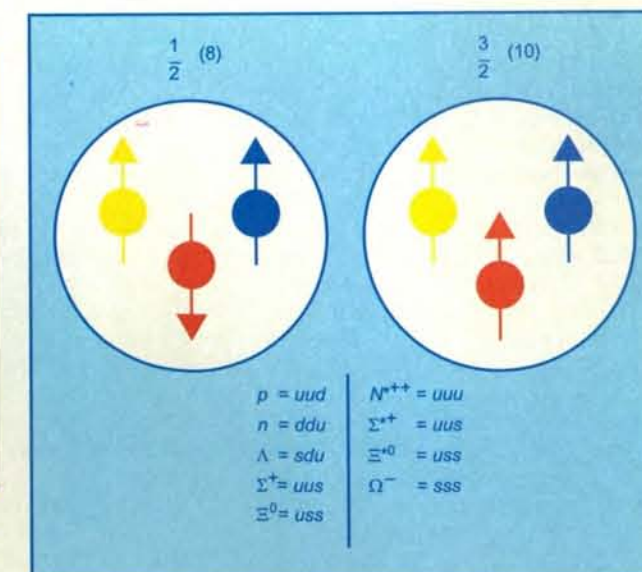
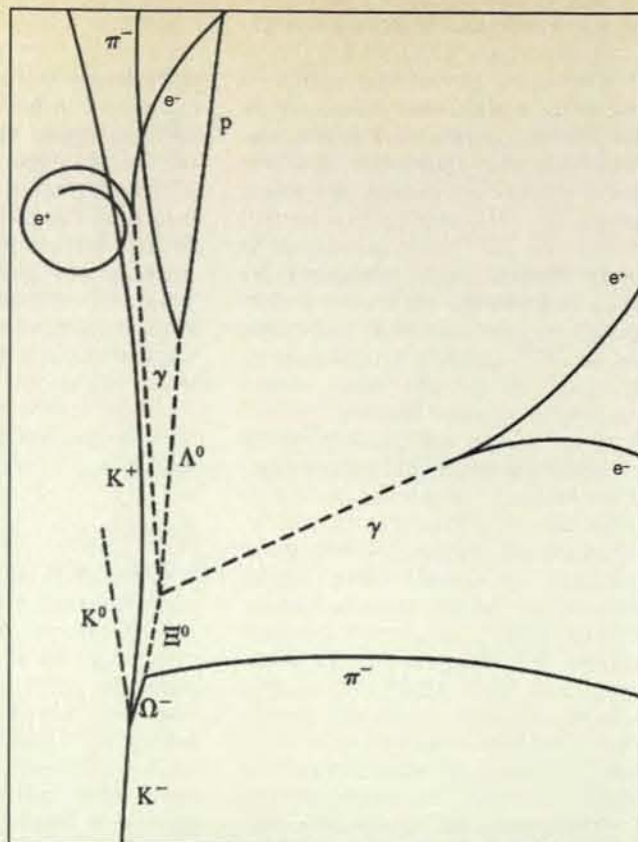
I modelli degli adroni

Come detto all'inizio, un modello semplice permise di capire, ancor prima che nascesse la QCD, diverse proprietà degli adroni, sia mesoni che barioni. Chi scrive (che lo propose all'inizio del 1965) e altri, tra i quali in particolare Richard Dalitz, lavorarono a lungo su di esso. Ci si meravigliò che un modello così semplice funzionasse bene; e la nascita della QCD, con la sua complicazione matematica, non contribuì certo a chiarirne il motivo. Il modello è chiamato «modello a quark non relativistico» (NRQM) ed è usato abitualmente; in inglese è detto anche modello «naive», proprio per la sua semplicità. Esso calcola le proprietà degli adroni come se, al loro interno, i quark si muovessero lentamente rispetto alla velocità della luce, come avviene per i nucleoni all'interno di un nucleo. Così, ad esempio, le proprietà del protone si calcolano in modo simile a quelle di un nucleo di ^{3}He fatto di quark (due u e un d) e le proprietà di un neutrone come quelle di un nucleo di ^{3}H composto di due quark d e uno u. Il fatto che il modello funzioni significa che approssima bene le soluzioni esatte. Solo oggi, però, si è capito perché funziona e, in quanto segue, cercherò di darne un'idea.

Per chiarire che cosa vuol dire calcolare le proprietà degli adroni mi riferirò alle masse dei barioni, indicate nella figura in alto a destra nella pagina a fronte accanto a ciascuno di essi. Un calcolo con la QCD, se fossimo capaci di farlo, darebbe queste masse e, in generale, le masse di ogni adrone (in ogni stato).

Come ripeto, le equazioni della QCD, però, sono troppo complicate. Ma, per i barioni indicati nella figura in alto nella pagina a fronte, la QCD (o qualunque teoria simile) permette di scrivere la massa di ciascuno come combinazione di certe quantità. Senza entrare in troppi dettagli, dirò solo che la massa M_B di ogni barione B si può esprimere («parametrizzare») in termini dei contributi dovuti ai quark contenuti nei barioni (*si veda l'illustrazione in basso a destra in questa pagina*) nella forma seguente:

$$M_B = c_0+c_1A_1+c_2A_2+c_3A_3+c_4A_4+c_5A_5+c_6A_6+c_7A_7,$$

Schema di costituzione dei mesoni (*a sinistra*). Un cerchio vuoto indica un quark e uno pieno un antiquark, mentre le frecce rappresentano l'orientazione dei loro spin. Sono indicate le composizioni di alcuni mesoni: il π^+, per esempio, così come il ρ^+, sono costituiti da un quark u e un antiquark d. Essi differiscono per l'orientazione relativa degli spin del quark e dell'antiquark che li compongono. A destra, schema di costituzione dei barioni. Gli antibarioni avrebbero tre antiquark al posto di tre quark. Si sono colorati i quark per ricordare che essi possiedono la caratteristica denominata «colore»; in un barione i colori dei tre quark devono essere diversi, come qui indicato. Come esempio sono rappresentate le composizioni di diversi barioni. In aggiunta ai quark indicati, un barione (e lo stesso dicasi per i mesoni) «contiene» una infinità di coppie quark-antiquark, nonché di gluoni. Le formule discusse nel testo tengono conto di questo fatto.

A sinistra, la fotografia in camera a bolle del primo Ω^-, rivelato nel 1964 per opera di un gruppo di fisici del laboratorio di Brookhaven. A destra, l'analisi dell'evento, che corrisponde alla sequenza di decadimenti: $\Omega^- \rightarrow \Xi^0 + \pi^-$; $\Xi^0 \rightarrow \Lambda^0 + \pi^0$; $\Lambda^0 \rightarrow p + \pi^-$; $\pi^0 \rightarrow 2\gamma$. Le particelle neutre non sono visibili in camera a bolle (non ionizzano e quindi non lasciano traccia), pertanto le linee tratteggiate a destra non esistono in effetti nella figura a sinistra. Esse sono state ricostruite usando in ogni decadimento la conservazione dell'energia e dell'impulso. Si noti anche che ambedue i γ provenienti dal decadimento del π^0 si sono materializzati in una coppia e^+, e^- nella regione visibile della camera a bolle: una circostanza, questa, assai fortunata.

dove c_0... c_7 sono otto coefficienti e A_n sono certe grandezze. Nel ricavare dalla QCD questa formula *non si fa nessuna approssimazione*. Si usano solo proprietà generali della QCD (ogni teoria con le stesse proprietà generali porta alla stessa formula). Gli A_n sono quantità di cui conosciamo il valore, diverso per ogni barione. Il termine A_1 nasce, riferendoci alla figura in basso a destra a pagina 77, dalla somma dei contributi dovuti a ciascun quark; i termini A_2, A_3, A_4 coinvolgono due quark alla volta, mentre A_5, A_6, A_7 coinvolgono tutt'e tre i quark. Se fossimo capaci di fare il calcolo, la QCD ci fornirebbe gli otto coefficienti c_n e, come ho già detto, se la QCD va bene, si otterrebbero le masse dei barioni date nella figura in alto a pagina 76. Non sappiamo far questo, ma possiamo fare il viceversa, cioè ricavare dalle masse gli otto coefficienti c_n. Poiché otto sono i barioni e otto i coefficienti c_n, ognuno di questi è univocamente determinato. Se la QCD è giusta, questi valori sono quelli a cui la QCD porterebbe se si sapesse fare il calcolo. I valori dei c_n sono riportati nella tabella in basso nella pagina a fronte.

Si capisce adesso perché il modello a quark non relativistico funziona. Due sono le ragioni: in primo luogo la formula scritta, che si ricava dalla QCD esattamente, ha la stessa struttura di quella tipica del modello a quark non relativistico, nonostante il fatto che, nel dedurla dalla QCD, non si facciano approssimazioni e la «bassa velocità» non c'entri. In secondo luogo, il modello non relativistico porta a una formula identica a quella scritta, ma con solo i primi tre (o quattro) degli otto termini indicati sopra; cioè nel modello le masse sono date dalla formula priva della parte con i coefficienti da c_4 a c_7. In conclusione, il modello dà per le masse dei barioni la stessa formula della QCD in cui mancano però i termini più piccoli, il che spiega appunto perché il modello funziona. Il fatto che i c_n decrescano al crescere di n, come indicato nella tabella in basso nella pagina a fronte, spiega anche come mai certe relazioni tra le masse dei barioni siano verificate meglio del previsto; e conduce a una nuova relazione verificata all'uno per mille.

Un'ultima osservazione: la formula per le masse, si è detto, è una conseguenza esatta della QCD. Ma nella QCD (e in ogni teoria simile) lo stato di un barione non è costituito solo di tre quark, bensì di tre quark più un numero infinito di coppie quark-antiquark di ogni tipo e colore più un numero indefinito di gluoni. L'esistenza di queste coppie e dei gluoni è una conseguenza generale della meccanica quantistica. Esse «si vedono» se si fa una «fotografia» del protone in tempi dell'ordine del milionesimo di miliardesimo di miliardesimo di secondo, che è quanto avviene in un urto ad alta energia. Tali «fotografie» si analizzano bene col cosiddetto «modello a partoni» e con la QCD. Grosso modo, le cose vanno come se ciascuno dei tre quark della figura in basso a destra a pagina 77 fosse a sua volta costituito da un quark nonché dalle coppie quark-antiquark e dai gluoni ora detti. Sorge la domanda se la formula scritta sopra tenga conto di tutto questo. La risposta è affermativa (altrimenti la formula non sarebbe certo esatta). È proprio la complicazione dovuta alle coppie quark-antiquark e ai gluoni a rendere matematicamente difficile il calcolo dei coefficienti c_n con la QCD; ma la formula scritta è esatta (le difficoltà sono tutte contenute nel calcolo dei c_n) e la sua deduzione chiarisce il collegamento tra il modello non relativistico e quello a partoni.

Parametrizzazione e modelli

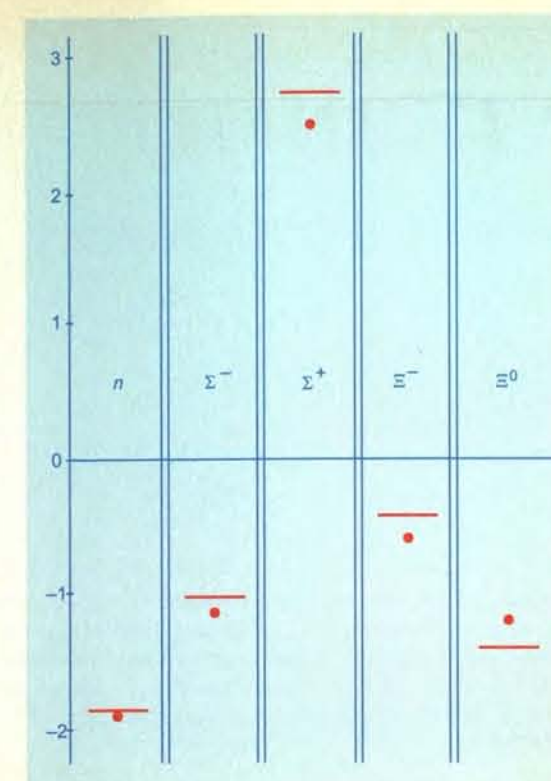
L'utilità della formula scritta prima si basa sul fatto che i c_n decrescono nel modo indicato. Se ciò non fosse vero, la formula varrebbe poco. Essa esprimerebbe otto masse con otto parametri, il che, ovviamente, è sempre possibile. La diminuzione dei c_n al crescere di n e il fatto che, usando solo i primi c_n, si ha una descrizione approssimata ma buona sono gli elementi fondamentale del giuoco.

Questo spiega altre cose. In primo luogo si capisce che, se ci si contenta di descrivere un certo fenomeno (le masse dei barioni, per esempio), modelli diversi tra loro e dal modello non relativistico possono funzionare tutti abbastanza bene. Basta che il modello proposto contenga un certo numero di parametri liberi (cioè sia abbastanza flessibile) e riprodurrà, scegliendo opportunamente i parametri stessi, i primi tre o quattro coefficienti nella formula scritta prima. Ecco che il modello funzionerà ragionevolmente bene per le masse. In effetti esiste attualmente una varietà di modelli per descrivere gli adroni, quello non relativistico e altri (di vario genere) in cui il moto dei quark è trattato come relativistico estremo, ovvero in cui i quark si muovono con velocità vicina a quella della luce. Quanto detto indica che non vale la pena di leticarsi su quale sia il modello più idoneo, se ci si limita a una sola categoria di fatti, come le masse. Semmai ci si deve chiedere qual è il più semplice e il più adatto a trattare un gran numero di fatti.

I momenti magnetici dei barioni

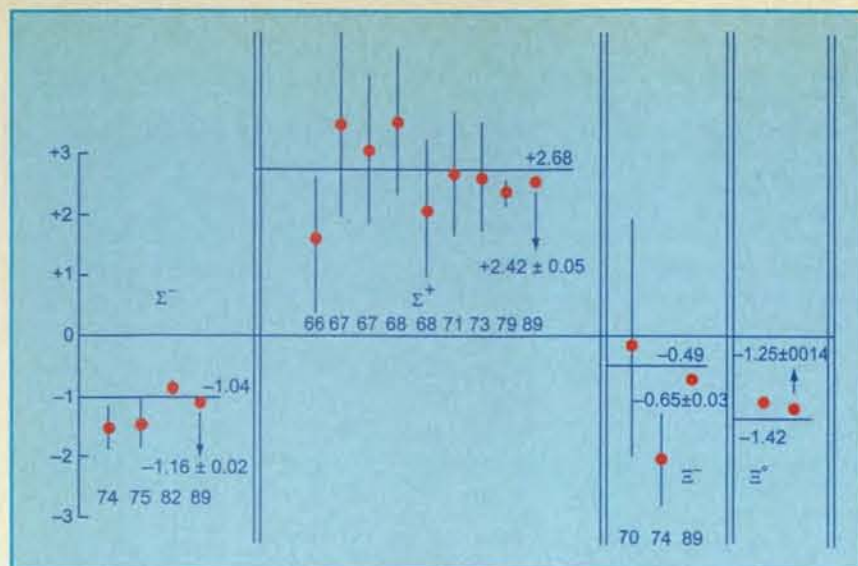
Tra le varie proprietà dei barioni, un'altra importante è data dai «momenti magnetici». Dal punto di vista magnetico ogni barione è simile a un aghetto magnetizzato; il momento magnetico è la magnetizzazione dell'aghetto (ossia del barione). A sua volta ogni quark in un barione ha un momento magnetico, cioè è anch'esso un aghetto magnetizzato.

Siamo in grado di misurare il momento magnetico di un barione; abbiamo buone misure di quelli di p, n, Λ, $\Sigma^{\pm}$, $\Xi^{0,-}$, Ω^-. (Tra i barioni elencati, quelli strani vivono circa un decimo di miliardesimo di secondo; quindi le misure non sono facili, ma hanno ormai una precisione di pochi per cento.) Il momento magnetico di un barione dipende dai momenti magnetici dei quark di cui il barione è composto e dal loro moto; lo si può calcolare, in linea di principio, con la QCD ma, come per le masse, il calcolo è troppo complicato. Tuttavia si può ottenere la formula più generale a cui la QCD conduce per i momenti magnetici dei barioni (cominciamo con quelli di spin 1/2, tutti quelli sopra elencati tranne l'Ω).

La QCD prevede che la formula generale sia una somma di termini simile a quella scritta prima per le masse. Ovviamente i termini sono del tutto diversi dai precedenti; inoltre questa volta la somma contiene molti addendi, ma omettendo quelli trascurabili ne rimangono sette, proprio il numero dei momenti magnetici misurati. Anche in questo caso, come per le masse (dove peraltro nella formula non si era omesso nulla), i dati sperimentali consentono di determinare i coefficienti, che chiamiamo questa volta b_n. Di nuovo si trova che la formula proveniente dalla QCD ha la stessa struttura di quella proposta dal modello non relativistico. Inoltre, tra i sette coefficienti b_n, quelli più significativi sono i primi due (b_1 e b_2); gli altri cinque contano assai meno (al più il 14 per cento di b_1). Usando solo i primi due coefficienti (dei sette anzidetti) l'accordo tra le previsioni e i momenti magnetici calcolati è piuttosto buono (*si veda l'illustrazione in alto in questa pagina*). Come per le masse, i termini dominanti b_1 e b_2 coinvolgono un quark alla volta, sono cioè additivi; gli altri termini, cui contribuiscono insieme due o tre quark, pesano solo il 14 per cento di b_1, come detto.

I momenti magnetici previsti dal modello a quark non relativistico usando soltanto i due parametri b_1 e b_2. I valori previsti (in certe unità di misura) sono indicati con linee orizzontali; i valori misurati sono indicati con punti. Si sono riportati solo i momenti magnetici di n, Σ^+, Σ^-, Ξ^0, Ξ^- perché quelli di p e Λ sono stati usati per determinare i coefficienti. Il momento magnetico di Ω^- (una cui misura accurata è stata possibile solo l'anno scorso) è discusso nel testo; per esso il valore misurato è $-2{,}024 \pm 0{,}056$, mentre quello previsto solo con i parametri b_1 e b_2 usati per ottenere i momenti magnetici riportati in figura è $-1{,}84$.

I COEFFICIENTI C_n

C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
= 1076	= 192	= 45.6	= −14	= 5	= −16 ± 1.4	= −1.1 ± 0.7	= 4 ± 3

I valori dei coefficienti c_n nella formula del testo. All'aumentare di n i coefficienti diminuiscono abbastanza rapidamente; è questo fatto che rende utile la formula.

I valori misurati dei momenti magnetici di Σ^+, Σ^-, Ξ^0, Ξ^- in funzione dell'anno della misura. Chiaramente la precisione delle misure è andata aumentando nel corso degli anni fino ad arrivare ai valori attuali attorno al 1989, sebbene si siano avuti alcuni miglioramenti anche più recentemente. (Il caso di Ω^- è un caso a sé: il risultato citato si è ottenuto nel 1995.) I livelli orizzontali sono quelli del modello a quark non relativistico già indicati nella figura a pagina 79.

Il modello a quark non relativistico, nella forma originaria, calcola i momenti magnetici come somma dei momenti magnetici dei quark contenuti in ciascun barione. Quindi esso è un'approssimazione alla QCD limitata ai termini b_1 e b_2; poiché essi dominano, si capisce che il modello funzioni.

In generale la migliore verifica che, nell'interpretare un fenomeno, siamo sulla strada buona si ha applicando la stessa spiegazione a misure eseguite successivamente: i momenti magnetici sono un buon esempio. Un anno fa è stato misurato il momento magnetico di Ω^- ottenendo, con un'incertezza di meno del 3 per cento, il valore –2,02 (nelle stesse unità in cui quello del protone vale 2,79). Il modello non relativistico prevede un valore di –1,84, con i soli due termini b_1 e b_2 che già spiegavano gli altri momenti magnetici; dunque si ha una verifica del modello (all'8 per cento). Di nuovo, ogni altro modello con due parametri capaci di produrre i termini b_1 e b_2 porta, naturalmente, agli stessi risultati.

Vale ancora una volta la conclusione che modelli anche molto diversi possono andar bene, se ci si contenta di descrivere un numero limitato di fenomeni e se il modello proposto ha il numero di parametri necessario.

Ma, naturalmente, modelli che si prefiggano di spiegare le masse e i momenti magnetici senza disporre del numero di parametri minimo (cioè che non riproducano i primi tre o quattro termini della formula per le masse e i primi due termini di quella analoga per i momenti magnetici) non possono andar bene. Certi modelli sono troppo ambiziosi; vorrebbero sostituirsi alla QCD senza esserle matematicamente equivalenti e senza considerare che, come primo requisito, devono approssimare le formule discusse prima in modo soddisfacente, cioè contenere, per ciascuna classe di fenomeni che si analizza, il numero di parametri appropriato.

In conclusione, oggi abbiamo una spiegazione soddisfacente del perché il modello non relativistico funziona e si è capito anche che l'espressione «non relativistico» è fuorviante (meglio sarebbe «quasi additivo» per ricordare che nella formula soprascritta i primi termini, additivi nei quark, sono i più importanti). Si è anche capito come tale modello sia collegato ad altre descrizioni adatte a trattare urti ad alta energia. Anche se restano aperte le grosse questioni del confinamento e, specialmente, della corrispondenza tra leptoni e quark, molti aspetti della struttura degli adroni stanno diventando trattabili come lo erano quelli della struttura dei nuclei quarant'anni fa e degli atomi settant'anni fa. Quale sarà il prossimo passo?

Il futuro è imprevedibile; questo va sempre sottolineato. Ma non è per nulla chiaro come la situazione evolverà. Per esplorare, per esempio, l'interno del protone su una scala sempre più fine, occorre bombardare il protone con proiettili di energia via via più alta e utilizzare solo i dati provenienti da una frazione degli urti sempre più piccola. Queste due circostanze insieme hanno la conseguenza seguente: supponiamo di voler costruire un acceleratore di particelle per effettuare misure della struttura del protone con «finezza» (potere risolutivo) dieci volte maggiore di quella ottenibile con un altro acceleratore oggi disponibile. La spesa di energia che occorre preventivare per la misura, qualunque sia il tipo di acceleratore, comunque perfezionato esso sia, risulta almeno mille volte maggiore. Oggi il costo dell'energia elettrica consumata da un grosso acceleratore è di qualche decina di miliardi di lire all'anno; perciò - a parità di rendimento dell'acceleratore - migliorare il potere risolutivo di un fattore dieci comporterebbe bollette di decine di migliaia di miliardi. Può sembrar curioso, ma è così: misurare distanze sempre più piccole costa energie sempre maggiori. Il limite alla misura delle piccole distanze, attualmente, sembra esser dato dalla quantità di petrolio esistente sulla Terra (o, meglio, da quella che i nostri condomini su questo pianeta pensano opportuno dedicare a tale questione).

GIACOMO MORPURGO si è laureato in fisica a Roma nel 1948 e ha vinto nel 1957 un concorso a cattedre di fisica teorica. Ordinario di tale disciplina e poi di struttura della materia, oggi insegna fisica delle particelle all'Università di Genova. È stato membro del Comitato per le direttive scientifiche del CERN ed è socio corrispondente dell'Accademia dei Lincei.

KOKKEDEE J. J., *The Quark Model*, Benjamin, New York, 1968.

MORPURGO G., *Introduzione alla fisica delle particelle*, Zanichelli, Bologna, 1987.

DILLON G. e MORPURGO G., *The Relation of Constituent Quark Models to QCD: Why Several Simple Models «Work so Well»* in «Physical Review D», 53, 1996, p. 3754 e lavori ivi citati.

Le immagini di pagina 75 e di pagina 78 sono tratte dal volume *Dai quark al cosmo* di Leon M. Lederman e David N. Schramm, Zanichelli, Bologna, 1991.

La madre di tutte le estinzioni

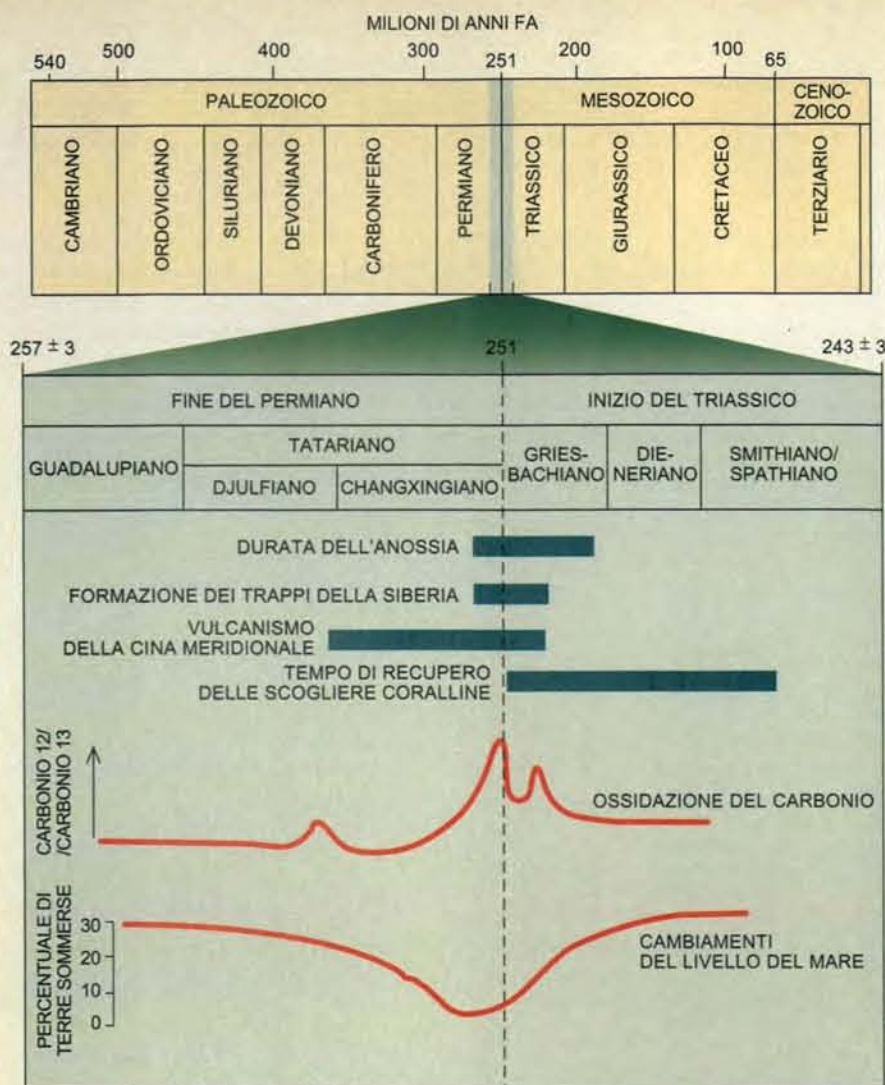
La più grave decimazione di specie della storia terrestre avvenne circa 250 milioni di anni fa; l'estinzione di massa del Permiano superiore diede una svolta fondamentale allo sviluppo della vita sul nostro pianeta

di Douglas H. Erwin

La storia della vita sulla Terra è piena di catastrofi di varia entità. Di esse, quella che si è assicurata gran parte delle attenzioni del pubblico e degli studiosi è l'estinzione dei dinosauri e di altri organismi, avvenuta 65 milioni di anni fa - al passaggio tra il Cretaceo e il Terziario - che fece sparire metà di tutte le specie allora esistenti. Ma nonostante la sua gravità, quella devastazione impallidisce al cospetto del più grande disastro mai avvenuto: l'estinzione di massa della fine del Permiano, circa 250 milioni di anni fa. Chiamata scherzosamente dai paleontologi «la madre di tutte le estinzioni di massa» (con buona pace di Saddam Hussein), essa reclamò un tributo impressionante. Circa il 90 per cento di tutte le specie presenti negli oceani sparì durante gli ultimi milioni di anni del Permiano; sulle terre emerse si ebbe la scomparsa di oltre due terzi delle famiglie di rettili e anfibi. Anche gli insetti non scamparono alla strage: anzi, fu questa la sola estinzione di massa che essi abbiano mai subìto: il 30 per cento degli ordini di insetti cessò di esistere.

Ma dalle catastrofi emergono anche nuove opportunità. Per parecchi milioni di anni prima dell'evento del Permiano superiore, i mari poco profondi erano stati dominati soprattutto da forme di vita immobili. Gli animali marini in gran parte poggiavano sul fondo o erano ancorati a esso per mezzo di fusti e si procuravano il cibo filtrandolo dall'acqua oppure aspettando pazientemente la preda. All'indomani dell'estinzione, molti gruppi un tempo secondari - predatori imparentati con gli attuali pesci, cefalopodi, gasteropodi e crostacei - ebbero la possibilità di espandersi, e apparvero linee evolutive completamente nuove. Questa riorganizzazione ecologica fu così drastica da costituire una demarcazione fondamentale nella storia della vita: non solo essa separa il Permiano dal Triassico, ma stabilisce anche il termine dell'Era paleozoica e l'inizio di quella mesozoica.

Negli ultimissimi anni, nuovi emozionanti dati sulle cause e sulle conseguenze dell'estin-

Alla fine del Permiano, circa 250 milioni di anni fa, un concorso di eventi catastrofici spazzò via la maggior parte delle forme di vita presenti sulla Terra. Le scogliere coralline e le comunità di mare poco profondo (costituite da coralli, gigli di mare, foraminiferi e altri organismi) furono colpite in modo particolarmente duro. Sulla terraferma si verificò la scomparsa di oltre due terzi dei rettili e degli anfibi e di circa un terzo degli insetti. Un incremento delle spore fungine fa pensare che anche le piante abbiano sofferto gravi perdite.

Il degrado dell'ambiente nel Permiano superiore è testimoniato da dati geochimici e fossili. All'epoca si verificò un picco nell'ossidazione del carbonio, iniziò un abbassamento del livello marino ed ebbe luogo un intenso vulcanismo nelle attuali Siberia e Cina. Alcuni strati di acqua negli oceani potrebbero anche essere divenuti anossici. Le scogliere coralline non si ripresero del tutto fino alla metà del Triassico.

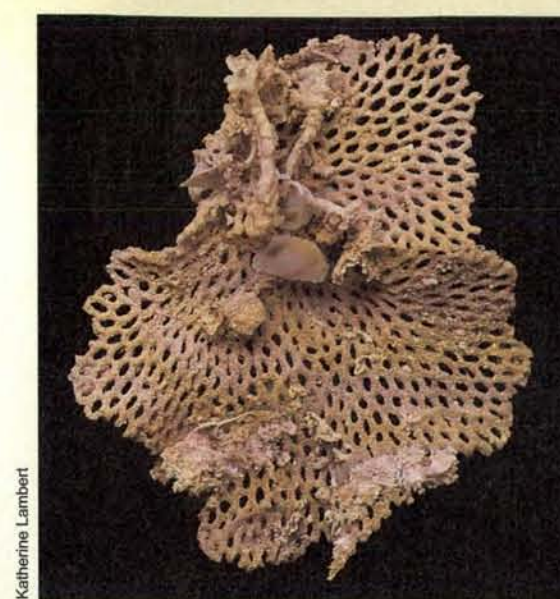
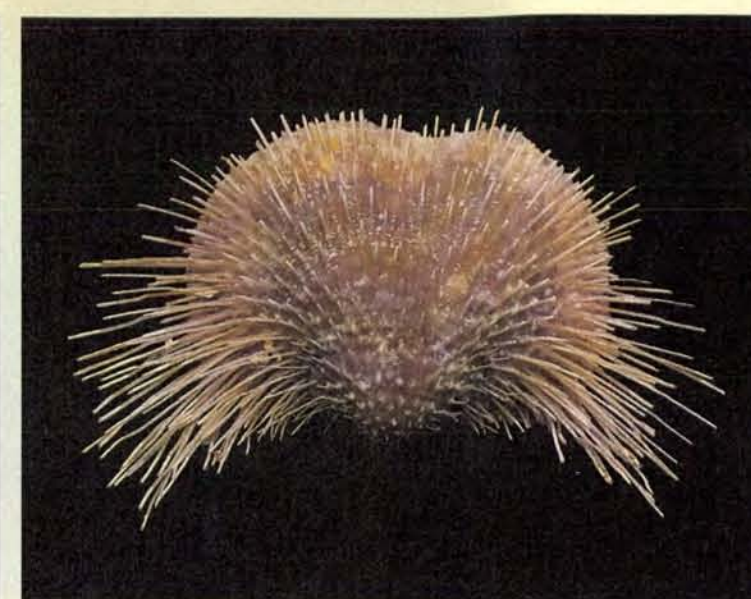
zione di massa del Permiano superiore sono giunti da quasi ogni ramo delle scienze della Terra. Alcune di queste scoperte includono studi di dettaglio dei rapidi cambiamenti nel chimismo degli oceani, una più completa documentazione dell'andamento delle estinzioni e nuove analisi che mostrano come al limite permo-triassico si siano verificate imponenti eruzioni vulcaniche (si veda l'articolo *Le grandi province magmatiche* di Millard F. Coffin e Olav Eldholm in «Le Scienze» n. 304, dicembre 1993).

Le mie ricerche nell'ultimo decennio sono state ispirate dalla curiosità riguardo agli eventi che determinano l'andamento dell'evoluzione. In quale misura, rispetto alle tendenze adattative a lungo termine, le estinzioni di massa contribuiscono all'evoluzione di un gruppo? Per esempio, i ricci di mare sono ubiquitari negli oceani attuali, ma erano relativamente poco comuni nel Permiano. Conosciamo un solo genere, *Miocidaris*, certamente sopravvissuto all'estinzione. Si trattò di un puro caso, oppure era meglio adattato di altri? E oggi i ricci di mare avrebbero caratteri diversi se non fosse avvenuta l'estinzione del Permiano superiore?

Per risolvere tali questioni, dobbiamo raccogliere qualche elemento in più sulle cause della catastrofe e su ciò che fece la differenza tra le specie sopravvissute e quelle scomparse. Le fonti essenziali per questo tipo di informazioni sono gli strati sedimentari e i fossili in essi contenuti. Purtroppo i campioni del Permiano superiore e del Triassico inferiore sono notoriamente molto rari. La documentazione fossile di questa fase è scarsa a causa della cattiva conservazione, della carenza di rocce da campionare e di problemi di altra natura. Un abbassamento generalizzato del livello marino durante il Permiano superiore ha limitato l'entità dei depositi sedimentari marini, e d'altronde i geologi hanno potuto raggiungere solo con molta difficoltà le aree in cui si trovano gli affioramenti migliori (soprattutto nella Cina meridionale).

Si è rivelato quindi assai difficile accertare con quale rapidità si siano verificate le estinzioni o se vi siano state variazioni regionali. Alcuni organismi, specialmente quelli più sensibili ai mutamenti ambientali, scomparvero rapidamente, come dimostrato da Erik Flügel e colleghi dell'Università di Erlangen, i quali arrivarono a questa conclusione dopo avere esaminato scogliere coralline nella Cina meridionale e in Grecia. Altri riscontri indicano estinzioni più graduali: per esempio, nello studio della incredibilmente diversificata e splendidamente conservata fauna degli affioramenti calcarei del Texas occidentale, del New Mexico e dell'Arizona, ho potuto osservare che molti gasteropodi cominciarono a estinguersi poco dopo la metà del Permiano, ben prima quindi del principale episodio di estinzione.

Lo studio intensivo di alcuni strati limite di grande importanza scoperti recentemente in Italia, Austria e Cina meridionale si è dimostrato illuminante, in quanto indica che la durata dell'estinzione fu più breve di quanto si sia ritenuto in passato; ciò implica che si siano instaurate improvvisamente condizioni ambientali calamitose. Solo alcuni anni fa, ritenevo che il periodo di estinzione potesse essersi protratto per 5-10 milioni di anni; ora sembra di poter dire che l'episodio finale sia durato meno di un milione di anni. Steven M. Stanley della Johns Hopkins University teorizza che l'estinzione possa essere consistita di due episodi distinti, l'uno avvenuto alla fine del Permiano medio e l'altro alla fine del Permiano superiore. Jin Yugan dell'Istituto di geologia e paleontologia di Nanchino, Samuel A. Bowring del Massachusetts Institute of Technology e io stiamo collaborando a un progetto per la datazione di depositi di ceneri vulcaniche nella Cina meridionale che dovrebbe presto fornirci una migliore stima della durata dell'estinzione. In ogni caso questa sembra essersi verificata all'incirca con la stessa rapidità di molte altre estinzioni di massa.

In questo intervallo geologicamente breve, gli oceani del Permiano videro una complessa serie di eventi di vita e di morte. È difficile riuscire a quantificare l'entità delle estinzioni dal punto di vista tassonomico, ma si ritiene che, su scala globale, siano stati spazzati via

Gli organismi permiani documentati dai fossili comprendono i briozoi (*a sinistra*) e i brachiopodi, due *phyla* strettamente correlati che prima dell'estinzione avevano un ruolo dominante nella vita marina. I fossili illustrati sono di circa 40 millimetri.

il 49 per cento delle famiglie e il 72 per cento dei generi. Dal momento che le specie sono più difficili da identificare, è anche arduo valutare in quale percentuale si siano estinte, cosicché le stime proposte da vari ricercatori sono assai disparate. Yin Hongfu e colleghi della China University of Geosciences hanno esaminato lo strato di roccia che corrisponde al piano stratigrafico di Changxing, nella Cina meridionale. (Questo piano, insieme con quello di Djulfian, corrisponde all'ultima delle due suddivisioni del Permiano superiore; ogni piano prende il nome dalla località nella quale si trova la documentazione fossile più chiara.)

Yin e colleghi hanno riferito che, delle 476 specie di invertebrati del Permiano superiore, 435, ossia il 91 per cento, scomparvero. (Altre stime della percentuale di specie estinte globalmente oscillano tra l'80 e il 95 per cento, ma il limite inferiore di questo intervallo è probabilmente il più realistico.) Per confronto, l'evento verificatosi alla fine dell'Ordoviciano, 439 milioni di anni fa, eliminò il 57 per cento dei generi marini; l'estinzione al limite Cretaceo-Terziario, che fece sparire i dinosauri, si portò via il 47 per cento dei generi esistenti.

La grande estinzione del Permiano superiore colpì alcuni animali più duramente di altri. I gruppi che vivevano fissati al fondo marino e si nutrivano filtrando materiale organico dall'acqua subirono le conseguenze più gravi. Tra questi animali vi erano i coralli, i brachiopodi articolati (invertebrati dotati di guscio), alcuni briozoi (organismi filtranti raggruppati in colonie) e una varietà di echinodermi (i gigli di mare). Altri gruppi marini decimati includevano le ultime trilobiti, i foraminiferi di basso fondale (un tipo di zooplancton) e gli ammonoidi (lontani parenti del nautilo). Gasteropodi, lamellibranchi e nautiloidi ne uscirono abbastanza bene, lamentando la perdita di pochi gruppi. Il solo gruppo marino che fu davvero indifferente al caos crescente fu quello dei conodonti, cordati primitivi che fungono da importanti marcatori cronologici in quanto il loro apparato boccale si conserva facilmente.

Le cose non andavano molto meglio sulle terre emerse. Sia i vertebrati terrestri sia gli insetti subirono perdite sostanziali; tra i vertebrati scomparve il 78 per cento delle famiglie di rettili e il 67 per cento di quelle di anfibi. La rapidità con cui si verificarono queste estinzioni rimane oggetto di dibattito. Precedenti studi condotti sui magnifici fossili della regione del Karrù, in Sudafrica, avevano fatto pensare che il declino fosse avvenuto in parecchi milioni di anni, forse con un paio di picchi nel tasso di estinzione. Tuttavia, alcuni lavori recenti indicano un declino più rapido, analogo a quello con cui avvennero le estinzioni in ambiente marino.

L'estinzione di molti insetti segna una trasformazione di prima grandezza della fauna. Dei 27 ordini di insetti noti dal Permiano, otto si estinsero presso il limite permo-triassico, quattro furono gravemente decimati, ma si ripresero, e tre riuscirono appena a sopravvivere per poi estinguersi nel Triassico. Questo è il solo evento significativo di estinzione di insetti che sia mai stato identificato, e testimonia della severità di condizioni ambientali che caratterizzò quell'epoca.

Anche la flora terrestre soffrì, ma non è possibile dire in quale misura, in quanto i dati su cui si potrebbero fare affermazioni in proposito sono, per ora, tutt'altro che certi. Esaminando fossili di foglie in Australia, Greg J. Retallack dell'Università dell'Oregon ha dimostrato nel 1995 che le estinzioni di specie vegetali furono di gran lunga più drastiche di quanto non si ritenesse. Gli eventi di estinzione condussero a quel rapido mutamento dei tipi vegetali dominanti che si riscontra nella documentazione fossile australiana. (La scomparsa delle piante potrebbe anche avere contribuito all'estinzione di quegli insetti che se ne cibavano.)

La testimonianza fornita dai pollini e dalle spore riflette più accuratamente gli effetti sulle piante. Negli strati del Permiano superiore, il polline delle gimnosperme (piante legnose come le conifere) è quasi assente, e gli strati successivi ospitano solo cellule fungine e qualche altro tipo di resto organico. Nel 1995 Henk Visscher e colleghi dell'Università di Utrecht, nei Paesi Bassi, hanno scoperto che questa espansione dei funghi sembra avere avuto inizio in una fase tardiva del Permiano - specificamente nel piano di Changxing - prima di raggiungere l'acme al limite permo-triassico.

A quanto indica la documentazione fossile marina e terrestre, il Permiano superiore fu un'epoca nella quale quasi tutto andò storto, almeno dal punto di vista di una specie intenzionata a sopravvivere. Che cosa potrebbe avere causato questa carneficina di massa? La sola cosa che a quanto pare non avven-

Katherine Lambert

Tra le vittime dell'estinzione di massa del Permiano superiore si annovera la fauna della regione sudafricana del Karrù, dove è stato portato alla luce questo fossile, lungo 36 centimetri, di *Oudendon* (*a sinistra*), un terapside dalle caratteristiche affini a quelle dei mammiferi. L'ammonoide a destra, del diametro di nove centimetri, è stato rinvenuto nel Texas.

ne fu l'impatto di una meteorite, l'evento che invece con ogni probabilità uccise i dinosauri. Alla metà degli anni ottanta un gruppo di geologi dichiarò di avere trovato tracce quasi impercettibili di iridio - un indicatore critico di collisione - nello strato limite permo-triassico nella Cina meridionale. Ma a dispetto di molti tentativi, nessuno è riuscito a dare consistenza a queste asserzioni.

Non vi è comunque penuria di «individui sospetti». Una possibilità è il vulcanismo: un indizio chiave è costituito dai Trappi della Siberia, strati solidificati di antiche lave. I Trappi (nome derivato dal vocabolo svedese per «scale», che descrive l'aspetto a gradini di questi depositi) includono almeno 45 effusioni distinte, e variano in spessore da 400 a 3700 metri. Essi coprono almeno 1,5 milioni di chilometri cubi, e forse più, dal momento che potrebbero estendersi verso ovest al di sotto di rocce più giovani fino agli Urali. (Al confronto, l'eruzione del Pinatubo avvenuta nel 1991 fu una specie di sbuffo, che produsse soprattutto ceneri vulcaniche, ma non magma. Un termine di confronto più adeguato potrebbe essere l'eruzione del Laki, in Islanda, che nel 1783 produsse 15 chilometri cubi di lava.)

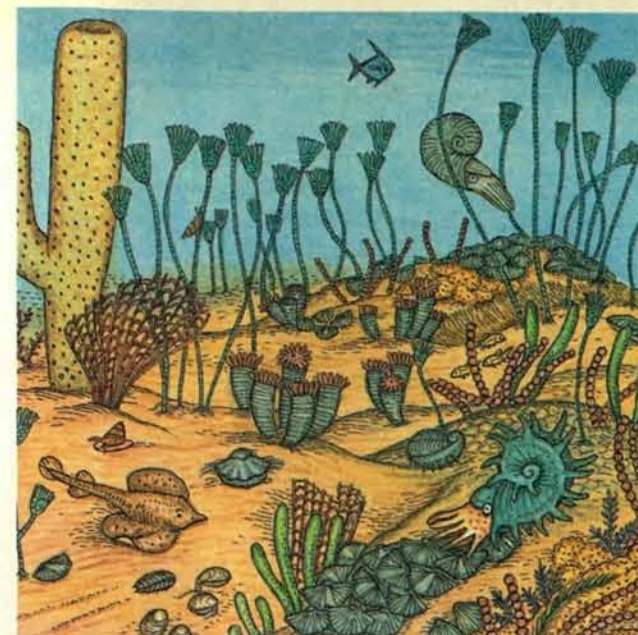
La recente datazione radiometrica indica che tutta la lava dei Trappi della Siberia venne eruttata in un arco di tempo inferiore a un milione di anni - forse solo 600 000 - a partire dal limite permo-triassico per proseguire nelle prime fasi del Triassico. Paul R. Renne del Berkeley Geochronology Center ha riscontrato che questa stima coincide con altre grandi eruzioni che provocarono la deposizione di ceneri vulcaniche nella Cina meridionale.

Ma un esteso vulcanismo potrebbe essere stato responsabile delle estinzioni del Permiano superiore? Le eruzioni hanno una varietà di effetti a breve termine, tra i quali il raffreddamento indotto dall'immissione nell'atmosfera di polveri e solfati, le piogge acide, gli incendi, la liberazione di elementi tossici e un incremento della radiazione ultravioletta dovuto all'impoverimento dell'ozono stratosferico. Inoltre, su tempi più lunghi, l'anidride carbonica emessa può provocare riscaldamento globale.

Per quanto attraente sia quest'ipotesi, uccidere il 90 per cento delle specie presenti negli oceani non è cosa facile. Di per sé il vulcanismo, pur rapido ed esteso come quello che produsse i Trappi della Siberia, non è all'altezza del compito. Thomas A. Vogel della Michigan State University e io abbiamo esaminato strati di ceneri vulcaniche depositati negli ultimi 100 milioni di anni. Queste eruzioni furono simili per entità a quelle che produssero la cenere trovata in Cina meridionale negli strati del Permiano superiore; eppure nessuno di questi eventi condizionò grandemente la diversità della vita a livello regionale o globale, sulla terraferma o negli oceani. Inoltre il danno ambientale prodotto da un'eruzione dipende da vari fattori. È difficile dedurre effetti del vulcanismo, come la quantità di solfati immessi nella stratosfera, relativi a eruzioni avvenute 250 milioni di anni fa. In conclusione, il vulcanismo può avere contribuito alle estinzioni, ma solo come parte di un processo più vasto.

La più interessante fra le nuove acquisizioni sull'estinzione di massa del Permiano superiore è data dalla geochimica. I cambiamenti forse più rilevanti dal punto di vista geochimico sono le alterazioni dei rapporti isotopici del carbonio nelle rocce (nello specifico, il rapporto fra carbonio 12 e carbonio 13). Questo fatto sembra indicare che nel Permiano superiore venisse sepolta più materia organica che non in precedenza.

Sebbene il seppellimento del carbonio ci dica qualcosa sui mutamenti geochimici coincidenti con l'estinzione del Permiano superiore, il messaggio non è del tutto chiaro. Un fattore può essere stato il repentino abbassamento del livello del mare. Nelle prime fasi del Permiano, i continenti si aggregarono a formare un solo supercontinente: il Pangea. Attorno alle piattaforme continentali prosperavano scogliere coralline e altre comunità di mare poco profondo. Verso la fine del Permiano, però, il livello del mare scese bruscamente. (Il fenomeno potrebbe essere stato causato da cambiamenti nella dinamica del mantello terrestre che avrebbero allargato i bacini oceanici.) L'abbassamento del mare distrusse gli habitat in prossimità delle coste. Dato che una notevole porzione della piattaforma continentale del Pangea si trovò esposta, i processi di erosione e di ossidazione di materia organica probabilmente si accentuarono. L'ossidazione ridusse i livelli di ossigeno nell'atmosfera e fece aumentare quelli di anidride carbonica, il che potrebbe avere reso il pianeta più umido e aumentato la temperatura media anche di due gradi Celsius.

Patricia J. Wynne

Patricia J. Wynne

PERMIANO

1 SPUGNA
2 CRINOIDE
3 BRACHIOPODE
4 NAUTILOIDE
5 SPUGNA
6 BRIOZOO
7 CORALLO
8 TRILOBITE
9 ALGA
10 GASTEROPODE
11 PESCE (*Janessa*)
12 PESCE (*Dorypterus*)

CRETACEO

1 CELACANTO
2 AMMONOIDE
3 BELEMNOIDE
4 GASTEROPODE
5 BIVALVI (RUDISTE)
6 RICCIO DI MARE
7 GRANCHIO
8 ALGA
9 PETTINE
10 PESCE (*Thrissops*)
11 PESCE (*Davichthys*)
12 STELLA DI MARE

La vita marina cambiò del tutto le sue caratteristiche in seguito all'estinzione di massa. A metà del Permiano (*a sinistra*) i mari contenevano per lo più animali immobili, alcuni pesci e poche trilobiti; ma all'inizio del Cretaceo (*a destra*) l'ambiente oceanico assomigliava ai mari attuali, con una massiccia presenza di animali mobili: bivalvi, gasteropodi, cefalopodi e pesci.

Quando il livello del mare risalì nuovamente, forse molte centinaia di migliaia di anni più tardi, si ebbero ulteriori distruzioni. Le acque dell'oceano invasero gli habitat vicini a quella che sarebbe diventata la nuova linea di costa, spazzando anche l'interno. Queste intrusioni provocarono senza dubbio la distruzione di molte comunità costiere.

Anche gli scarsi livelli di ossigeno atmosferico potrebbero avere esacerbato le condizioni ostili che già si stavano instaurando. Negli oceani poteva disciogliersi una minore quantità di ossigeno; la conseguente anossia potrebbe avere soffocato certe forme di vita marina. Indicazioni dell'esistenza di acque anossiche vengono dalle anomalie geochimiche. Alcuni studiosi hanno di recente esposto un'argomentazione interessante, sebbene non del tutto cogente, secondo la quale gli andamenti delle estinzioni da specie a specie rifletterebbero le capacità degli organismi di far fronte a condizioni di anossia.

In definitiva, tutte le possibilità finora menzionate potrebbero avere contribuito all'estinzione di massa. Nessuno di questi fattori, da solo, poteva causare un disastro di una simile entità, ma, per somma sfortuna di quelle meravigliose faune, tutti interagirono quasi simultaneamente.

Io ritengo che l'estinzione sia consistita di tre fasi. La prima ebbe inizio con l'abbassamento del livello marino intorno al Pangea, fenomeno che provocò distruzione di habitat, instabilità climatica ed eliminazione di molte specie con limitati areali di diffusione. Al procedere della regressione oceanica, ebbe inizio la seconda fase, caratterizzata da eruzioni vulcaniche e dall'immissione nell'atmosfera di grandi volumi di anidride carbonica, la quale accentuò l'instabilità climatica e avviò il collasso ecologico.

La terza fase, a cavallo tra Permiano e Triassico, cominciò con la risalita del livello marino e le conseguenti inondazioni da parte di acque verosimilmente anossiche. Questa fase distrusse gli habitat terrestri presso le coste e contribuì all'estinzione di molti *taxa* che avevano resistito fino a quel momento.

Le conseguenze dell'estinzione del Permiano superiore sono interessanti almeno quanto l'evento stesso. Dopo altre estinzioni di massa, la vita iniziò a riprendersi entro un milione di anni circa, ma in questo caso ci vollero forse cinque milioni di anni. (Vi è anche la possibilità che il recupero appaia più lungo di quanto fu in realtà a causa della cattiva conservazione dei fossili.)

Comunque, prescindendo dalla lunghezza dei tempi di recupero, la vita sulla Terra aveva subito drastici cambiamenti. Come notavo in precedenza, i

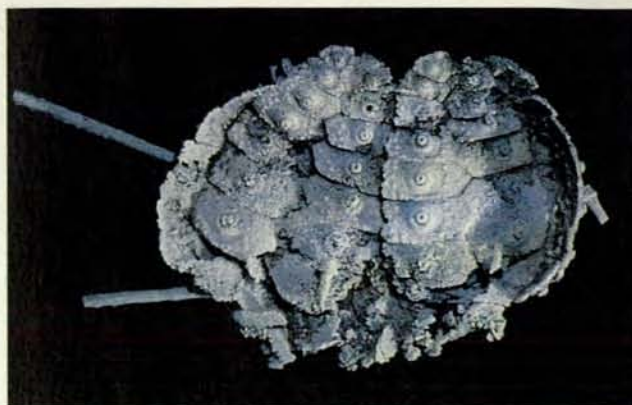
mari del Permiano erano stati dominati da animali per lo più immobili: brachiopodi, briozoi ed echinodermi. Questi organismi poggiavano sul fondo, filtrando l'acqua per estrarne il cibo o aspettando che la preda transitasse nei paraggi. Gli animali mobili - pesci, bivalvi, cefalopodi e gasteropodi - costituivano solo una piccola parte della comunità; restavano inoltre poche trilobiti.

Poco dopo l'estinzione, in quella che viene definita fase di sopravvivenza del Triassico inferiore, le poche specie restanti tendevano a essere abbondanti e molto diffuse. Le faune degli esordi del Triassico consistono di alcuni molluschi, ammonoidi e pochi gasteropodi. Il terapside erbivoro *Lystrosaurus*, progenitore dei mammiferi, era il vertebrato più comune sulla terraferma e si trovava in tutto il Pangea. Il mollusco *Claraia* prosperava negli oceani. Ma alla metà del Triassico, circa 25 milioni di anni più tardi, i ricci di mare e altri gruppi relativamente sensibili all'ambiente iniziarono a riemergere, ponendo le basi per il ritorno a condizioni marine più normali. Questi *taxa* redivivi cominciarono a rimpiazzare la fauna sopravvissuta.

In effetti, i mari di quell'epoca presentavano una certa somiglianza con gli oceani attuali. Predominavano animali dotati di mobilità, come i bivalvi, i gasteropodi e i granchi; emerse anche una maggiore diversità di cefalopodi e di altri predatori natanti.

Si moltiplicarono gli organismi scavatori, fenomeno che forse riflette una strategia di difesa dalla predazione. Si instaurò una «corsa agli armamenti» evolutiva tra predatore e preda, la quale produsse cambiamenti nelle strutture scheletriche. Ciò portò allo sviluppo, attraverso tutto il Mesozoico, di faune dotate di una maggiore quantità di tessuti molli rispetto agli animali dell'era precedente. Tali cambiamenti produssero ecosistemi più complessi e sofisticati: c'era sicuramente più da mangiare e anche una maggiore scelta nel menu.

I dettagli dei cambiamenti evolutivi verificatisi su terraferma nello stesso periodo sono alquanto incerti, dal momento che i fossili devono ancora essere campionati sistematicamente. Gli studi che ci si propone di condurre promettono di ampliare considerevolmente la nostra conoscenza delle estinzioni su terraferma. Per il momento sappiamo che molti gruppi di anfibi e rettili trovarono la fine, e anche che gli insetti modificarono notevolmente le loro caratteristiche, passando da una varietà di gruppi simili alle libellule - dotati di ali fissate nella posizione di volo e impossibili da ripiegare sul corpo - a gruppi che invece erano in grado di ripiegare le ali. Queste forme più nuove, che costituiscono il 98 per cento delle specie di insetti attuali, avevano anche fasi separate di sviluppo: larvale e adulto. Gli adattamenti potrebbero riflettere una capacità di sfruttare nuovi habitat e di fronteggiare forti oscillazioni stagionali e altre instabilità climatiche.

Douglas H. Erwin

Questo riccio di mare del Permiano, della lunghezza di quattro centimetri, appartiene all'unico genere di echinodermi - *Miocidaris* - che riuscì a sopravvivere all'estinzione di massa.

I cambiamenti che ebbero luogo fra gli insetti sollevano un problema più generale: si tratta di capire se le specie sopravvissute erano dotate di adattamenti specifici che permisero loro di sopravvivere oppure se questa sopravvivenza fu un fatto più accidentale. I fossili di *Claraia* si trovano in rocce che recano inequivocabilmente tracce di condizioni anossiche. Il gran numeri di esemplari ritrovati e la loro ampia distribuzione potrebbero indicare che questa specie era in grado di sopravvivere con poco ossigeno. Un altro esempio è *Miocidaris*, il solo echinoideo (riccio di mare) sicuramente sopravvissuto (sebbene sia assai probabile che anche un suo parente stretto vi sia riuscito). *Miocidaris* ha solo due colonne di piastre interambulacrali (per esprimersi in modo un po' improprio, le aree comprese tra i petaloidi sul guscio di un riccio di mare); gli altri echinoidei del Permiano potevano avere da una a otto colonne di piastre. Dal momento che *Miocidaris* fu l'unico genere a sopravvivere, la forma predominante di echinoideo passò da quella che aveva un numero fortemente variabile di colonne di piastre a quella che ne aveva solo due. Alcuni paleontologi hanno ipotizzato che lo scheletro di echinoideo fosse particolarmente robusto se composto da solo due colonne di piastre, e pertanto meglio adattato a resistere alla predazione nel mondo post-Permiano.

Sfortunatamente è quasi impossibile dire se sia stata proprio l'estinzione permo-triassica a permettere a talune caratteristiche di imporsi. Non si può escludere che gli echinoidei moderni avrebbero finito per sviluppare due colonne di piastre anche se l'estinzione del Permiano superiore non fosse mai avvenuta. La fauna sopravvissuta potrebbe semplicemente consistere di quei gruppi che prima dell'estinzione erano i più abbondanti e i più ampiamente distribuiti e che quindi avevano le migliori opportunità di sopravvivenza. Distinguere tra queste due possibilità si è rivelato assai arduo.

La sola cosa che possiamo affermare con sicurezza è che l'estinzione di massa del Permiano superiore, tra tutti gli eventi verificatisi dalla comparsa degli organismi complessi, fu quello con effetti più cospicui sulla storia della vita. Senza questo episodio, non vi è dubbio che la composizione delle attuali comunità biologiche della zona di marea avrebbe caratteristiche assai diverse.

DOUGLAS H. ERWIN è un paleobiologo che lavora presso il National Museum of Natural History della Smithsonian Institution; ha studiato rocce di età cambriana e permiana in Siberia, Cina, Terranova e in tutti gli Stati Uniti occidentali.

RAUP DAVID M., *Extinction: Bad Genes or Bad Luck?*, W. W. Norton, 1991.

SWEET WALTER C., YANG ZUNYI, DICKINS J. M. e YIN HONGFU (a cura), *Permo--Triassic Events in the Eastern Tethys: Stratigraphy, Classification, and Relations with the Western Tethys*, Cambridge University Press, 1992.

ERWIN DOUGLAS H., *The Great Paleozoic Crisis: Life and Death in the Permian*, Columbia University Press, 1993.

ERWIN DOUGLAS H., *The Permo-Triassic Extinction* in «Nature», 367, 20 gennaio 1994.

L'ANGOLO MATEMATICO

di Ian Stewart

Monopoli è un gioco equo?

Tutti hanno giocato a Monopoli. Immagino che pochi, però, abbiano pensato agli aspetti matematici del gioco. Eppure esistono interessanti costruzioni, chiamate catene di Markov, con le quali si possono descrivere le probabilità di vincere a Monopoli. Agli inizi del secolo, il matematico russo Andrej Andrejevič Markov inventò una teoria generale della probabilità. Tralascerò gran parte del suo lavoro, e non passerò neppure in rassegna tutte le regole del Monopoli: ciò che voglio fare è dimostrare che il gioco è equo.

Innanzitutto riassumiamo come si gioca. I giocatori lanciano a turno due dadi e la somma dei punti determina di quante caselle devono muovere lungo il percorso. Un giocatore che ottenga un punteggio doppio - per esempio due 1 - ha diritto a un altro lancio. Tutti i giocatori partono dalla casella contrassegnata con VIA!

Alcuni punteggi, come il 7, ricorrono più spesso di altri. Ci sono sei modi per ottenere un 7 (1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2, 6 + 1) tra le 36 possibili somme delle facce dei dadi. La probabilità che esca un 7, quindi, è 6/36, ossia 1/6. Seguono 6 e 8, ciascuno dei quali ha una probabilità pari a 5/36, poi 5 e 9, con una probabilità di 1/9; 4 e 10 hanno probabilità 1/12; 3 e 11 hanno probabilità 1/18. Infine, 2 e 12 hanno probabilità 1/36.

In base a questi valori sappiamo che, sulla distanza di numerose partite, il primo giocatore ha molte probabilità di finire sulla settima casella, IMPREVISTI. Se non esce un 7, il primo giocatore arriverà probabilmente su Bastioni Gran Sasso o Viale Monterosa e avrà così l'opportunità di assicurarsi una di queste proprietà. Se ne compra una, riduce la possibilità che gli altri giocatori hanno di comprare qualcosa al primo lancio di dadi.

I matematici vedono il tabellone del Monopoli come un cerchio in cui ciascuna casella è collegata alle sei successive.

Questo è indubbiamente uno dei motivi per cui gli inventori del gioco hanno messo vicino alla partenza proprietà di scarso valore. I costosi ma redditizi Viale dei Giardini e Parco delle Vittorie sono molto distanti, in un punto in cui è presumibile che le probabilità dei diversi giocatori siano pari. Ma è proprio così?

Per affrontare il problema mi servirò di un'altra semplificazione. Immaginiamo che i due dadi vengano lanciati non insieme ma separatamente. Ciascun giocatore può fare due mosse: una mossa «fantasma», in cui si trascura il punto di arrivo, e una mossa reale. Analogamente, adotteremo una rappresentazione matematica del tabellone del gioco.

Per convenienza, numeriamo le caselle da zero a 39. La casella 40 equivale alla zero, il VIA!, e si possono contare i numeri modulo 40: ciò significa che qualsiasi numero maggiore di 39 può essere sostituito dal resto della sua divisione per 40. Immaginiamo ora un solo giocatore che lanci ripetutamente un singolo dado e muova di conseguenza. Qual è la sua probabilità di finire su una data casella dopo un dato numero di lanci? La speranza è che al crescere del numero di lanci, la probabilità si avvicini a 1/40 per tutte le 40 caselle.

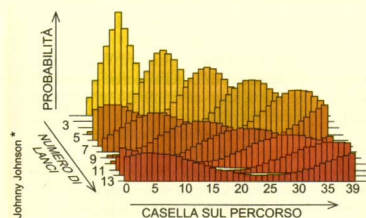
Il modo per individuare queste probabilità è vedere come la loro distribuzione «fluisce» nel tempo. Ciascuna distribuzione può essere rappresentata da una successione di 40 numeri, ognuno dei quali dà la probabilità di arrivare su una singola casella. All'inizio del gioco, il giocatore si trova sulla casella zero (VIA!) che ha probabilità 1. La distribuzione delle probabilità, quindi, è 1 seguito da 39 zeri. Dopo un lancio fantasma, la distribuzione diventa 0, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 0,..., 0: il giocatore ha probabilità pari a 1/6 di finire su una delle prime sei caselle e nessuna di raggiungere una casella successiva.

Si noti che la probabilità totale, 1 - che inizialmente era concentrata sulla casella zero -, si è ora divisa in sei parti uguali distribuendosi sulle caselle che distano da una a sei unità. Questo procedimento ha un valore generale: dopo ciascun lancio del dado, la probabilità relativa a una casella si divide per sei. Queste sei parti uguali fluiscono in senso orario su ciascuna delle successive sei caselle. Al lancio successivo, quindi, la probabilità di 1/6 della casella uno si distribuisce come segue: 0, 0, 1/36, 1/36, 1/36, 1/36, 1/36, 1/36, 0,..., 0. In modo analogo si ridistribuiscono le probabilità di 1/6 delle caselle da due a sei, spostate del corrispondente numero di passi.

La matrice magica di Markov

Sia M la matrice di transizione. Innanzitutto calcoliamo un insieme di 40 numeri, detti autovalori di M. Un numero m è un autovalore di M se si possono scrivere 40 numeri sui 40 vertici del reticolo in modo che, quando si suddivide ciascuno di essi in sei e lo si fa fluire lungo le sei linee che si dipartono in senso orario da quel vertice, i numeri risultanti sono grandi esattamente m volte i numeri con cui si è partiti. (Forse è più facile in notazione simbolica: $Mv = mv$ per un certo v.) Ma c'è una complicazione: non è detto che gli autovalori siano numeri reali compresi tra 0 e 1. Possono anche essere numeri complessi, esprimibili utilizzando il numero $i = \sqrt{-1}$.

La successione formata da questi 40 numeri è detta autovettore. Ora non ci resta che trovare l'autovalore più alto tra quelli calcolati. A quel punto la distribuzione delle probabilità sarà approssimata quanto si desidera dal corrispondente autovettore, una volta «normalizzato» in modo che i suoi valori diano per somma 1, come dovrebbe avvenire per le probabilità. (Questo passo significa semplicemente che si divide ciascun valore per il totale.)

A causa della simmetria rotazionale, non è in realtà difficile trovare gli autovalori e gli autovettori. Un autovettore ha tutti i 40 valori uguali a 1/40. Qual è il suo autovalore? Supponiamo che si inizi da questa distribuzione; dividiamo ciascun 1/40 in sei parti uguali di dimensione 1/240 che spostiamo lungo le loro sei linee in senso orario. Ciascun vertice riceve esattamente sei contributi: uno da ciascuno dei sei vertici precedenti. Si finisce quindi con 6 × 1/240, ossia 1/40. Questo è quanto dovrebbe fare un autovettore, e in questo caso l'autovalore è 1. Non vi rivelerò gli altri 39 autovalori, che hanno espressioni molto eleganti (ma forse solo per i matematici). L'autovalore seguente, in ordine di grandezza, ha un valore assoluto di 0,964. Quindi 1 è l'autovalore più grande, e il suo autovettore rappresenta effettivamente lo stato a lungo termine della distribuzione di probabilità.

Il grafico rappresenta la distribuzione di probabilità sulle 40 caselle e il modo in cui tale distribuzione varia a ciascun lancio di un dado. L'altezza di ogni barra dà la probabilità di finire sulla casella corrispondente. I grafici relativi ai lanci da 2 a 13 sono disposti l'uno davanti all'altro a partire dal fondo.

Infine sommiamo le probabilità attribuite a ogni casella. Per esempio, la casella sei riceve 1/36 dalle prime cinque successioni e 0 dall'ultima, quindi il totale è 5/36. Il risultato finale è 0, 0, 1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36, 0,..., 0. Questa distribuzione corrisponde a quanto ci aspettavamo dal lancio di due dadi. Al terzo lancio, moltiplichiamo ogni termine della nuova successione per 1/6, e attribuiamo il risultato alle caselle che distano di uno, due, tre, quattro, cinque e sei posti. Poi sommiamo i numeri per ciascuna casella.

È facile scrivere un programma per calcolatore che trovi una per una queste distribuzioni di probabilità. Nell'illustrazione di questa pagina si vede il risultato, che inizia con la distribuzione «triangolare» ottenuta al secondo lancio. A ciascun lancio successivo, il grafico delle probabilità si sposta di un passo in avanti nella figura. Si può vedere come, a ciascun passo, il picco di probabilità si sposti di diverse caselle verso destra. (Per la precisione, esso si muove in media di 3,5 caselle, pari al valore medio tra i numeri 1, 2, 3, 4, 5, 6.) Proseguendo la simulazione al calcolatore, si osserva che la forma triangolare finisce con l'appiattirsi e che tutti i valori diventano quasi uguali. Ma perché la simulazione segue questo andamento?

Per spiegarlo dobbiamo far ricorso alla teoria di Markov, che ci fornisce un metodo sistematico per ricostruire il flusso delle probabilità. Il metodo inizia con la stesura della cosiddetta matrice di transizione per il primo numero. La matrice, che chiameremo M, è una tabella quadrata con 40 righe e 40 colonne, ciascuna numerata da 0 a 39. Il valore inserito all'intersezione della riga r e della colonna c è la probabilità di passare, con una mossa, dalla casella r alla casella c. Il valore è 1/6 se $c = r + 1$, $r + 2$,..., $r + 6$ (modulo 40), e 0 negli altri casi. Segue poi un calcolo, che utilizza M, il cui risultato dimostra che effettivamente dopo numerosi lanci la probabilità si approssima a 1/40 per qualsiasi casella data. Con l'aiuto della teoria di Markov, possiamo quindi dimostrare che un gioco complesso come il Monopoli è equo, nel senso che - sul lungo periodo - nessuna casella ha maggiori o minori probabilità di essere raggiunta. Naturalmente il primo giocatore ha sempre un piccolo vantaggio, mitigato però dalla limitatezza della sua disponibilità finanziaria.

Per una spiacevole dimenticanza, nella foto pubblicata in questa rubrica a pagina 95 del numero di agosto non è stato indicato il nome del fotografo Ugo Allegri, di Brescia. Ce ne scusiamo con l'autore e con i lettori.